目　录

（上　册）

緒論　概率論的性质

1. 背　景

概率論是一門数学学科，它与几何学或分析力学等学科有很相似的目标。对每一門学科都必須仔細地去区分理論的三个方面：(a)形式邏輯的內容；(b)直观的背景；(c)应用。不按这三方面之間的固有关系去考虑这三方面，就不能正确估計其全部結构的特性和优点。

(a) 形式邏輯的內容. 公理化的数学只論及某些无定义的事物之間的关系。这一点可以用下象棋来作很好的說明。要想描述象棋，除了陈述一組下棋規則外，不可能給象棋下一个"定义"。尽管对棋子的通常的形式可以作某种程度的描写，但哪一个是"王"棋不一定总是很清楚的。棋盘和棋子是有用的，但沒有它們仍然可以下棋。重要的事情在于了解棋子的走法与作用，也就是要了解一組棋規。要問象棋中的一个"卒"或"王"的"定义"或"精确的本性"是什么，这是沒有意义的。与此类似，在几何学中我們也不去过問点和直綫到底是什么。点和直綫是无定义的概念，而几何公理規定它們之間的一些关系，例如，二点确定一直綫等等。这些就是規則，它們无异議地被承認。我們可以改变公理系統而去研究不同形式的几何，而且各种非欧几何的邏輯結构不依賴它們对現实的关系。物理学家曾研究，在不同于牛頓的万有引力定律时，物体如何运动，这种研究也是有意义的，即使牛頓的万有引力定律是正确的。

(b) 直观的背景. 与象棋对比，几何学和力学的公理反映了客观存在着的直观背景。事实上，几何的直观性是如此之強，以致它常常会跑在邏輯推理的前面。邏輯、直观与实际經驗三者之

間的相互依賴性达到怎样的高度，这是一个問題，我們不需要去討論。当然，人們的直观能力是可以鍛炼和发展的。例如，下象棋时，初学者小心翼翼地走棋，还要想一想各别的下棋規則；但是，有經驗的棋手在一瞥之下就能掌握复杂的情况，对他的直观往往不能用理由来解释。同样，数学直观随着經驗增加，人們有可能，例如說对于4維空間发展一种自然感覚。

甚至人类的集体直观能力也在进步着。牛頓的力場概念、超距作用的概念、馬克司威尔(Maxwell)关于电磁"波"的概念，起初都被認为是"不能設想的"和"违反直观的"。近代工艺学和每家都有一架无綫电，使得这些概念普及化而成为一些通常的語汇。同样，現在的学生不会体会到，当概率論还在萌芽的时候，与某些思維方式、偏見及其它困难的斗爭情形。現在，报紙报导着民意測驗的样本，統計的影响已經滲透到生活的各个方面。于是对于象这样的陈述："这个事件的机会是五分之三"，人人都有了直观的感覚。这虽然是很模糊，但此直观足够作为概率論入門的指南和背景。随着理論的发展，以及接触到一些更为微妙的应用，这个直观还要得到发展。

(c) 应用. 应用几何学和力学的概念时，要把它們与某些物理对象等同起来，但是这个結合过程是如此的灵活及变化无常，以致不能給出普遍的法則来。刚体这个概念是基本的而且是有用的概念，虽然沒有一个物理对象眞正具备它的条件。哪些物体能当作刚性的来处理，要視所处环境和所需要的近似度来决定。橡皮当然不是刚体，但是在討論汽車的运动的时候，許多教科书是把橡皮輪胎当作刚体来处理的。按照理論的目的，我們不管物質的原子結构，而有时把太阳当作一个連續物質的大球来处理，有时又把它当作一个質点来处理。

在应用中，抽象的数学模型只是当作工具来使用，而对同一实际場合可以采用不同的模型来描述。数学理論的应用方式不依賴于事先形成的意見；它是一个有目的的技术，依賴于經驗，而且随着經驗改变。值得研究这些技术的哲学分析，但它不屬于数学、

物理学或統計学范围以內． 因此，必須把关于概率基础的哲学从数学和統計学中分离出去，犹如关于直观空間概念的討論，現在已經从几何学中分离出去一样．

2. 程　　序

概率論的历史(一般数学史亦然)呈現出理論和应用的相互促进作用：理論的进展开辟了应用的新領域，反过来每一个新的应用又产生出新的理論問題和有成果的研究．目下概率論已应用到很多不同的領域，而我們所要求的是一个普遍性理論所具有的那种伸縮性，以便对广泛的种种需要提供恰当的工具． 因此我們必須反对下述企图(与趋向)： 把理論、它的术語以及它的武器建立得过分接近于某一特殊兴趣范围．我們所要作的絕非如此，而是按照在几何学和力学中已經証明为十分成功的方式来发展出一种数学理論．

我們先从扔錢币和擲骰子等最簡单的实驗出发，在这些場合，所有的陈述都有着明显的直观意义． 我們把这个直观性翻譯成一个抽象的模型，然后把这模型加以推广，使它能够适用于更复杂的場合．书中的实例是用来闡明几种模型的經驗背景和启发讀者的直观能力的，但理論本身还是数学性质的． 我們不拟力求解释概率的“眞正意义”，正如近代物理学家不致力于闡明质量和能的“实在意义”；几何学家不致力于解释点的性质那样． 我們将要做的，就是証明一些定理，并指出这些定理怎样应用．

最初，概率論的目的是描述有关机会游戏的經驗这一窄狭的領域，这时，主要目标只是要把某些概率計算出来．在开始的几章里，我們也还計算一些典型的概率，但是，要記住数值的概率不是理論的主要对象．其目的在于发現一般的規律并构造出滿意的理論模型．

概率对我們所起的作用正如质量在力学中所起的作用．在力学上，即使不知道个別行星的质量，不去思考实际上測量这些质量的方法，也照样可以討論行星体系的运动． 甚至虛构的行星体系

也可以作为一个有益的、有启发的研究对象．类似地，一些实际的且有用的概率模型涉及一些不能观察到的世事．例如，已經把几十亿元投資于自动电話交换．这些都是因为从簡单的概率論模型来考虑时，对其中各种可能的系統都需作比較．理論上的最优系統被建立起来了，而其余的系統也就不存在了．在保险业中，概率論用来計算毁灭概率，就是說用这理論来避免某些不期望发生的情形，因此，它应用到一些实际上观察不到的情形，即使連一个数值都不能得到时，概率論仍然是有效的和有用的．

3. "統計"概率

近代数学概率論的成就是有代价的，那就是把理論局限在"机会"的某一个特殊方面．概率的直观概念联系到归納推理，以及象下面的一些判断："保尔大概是一个幸运的人"；"这本书大概是一部失敗之作"；"費馬(Fermat)的猜測大概是不正确的"．这类判断受到哲学家和邏輯学家的注意，也是数学理論[1]的正确对象．然而必須这样了解，我們所关心的不是归納推理的型态，而是一种可以叫作是物理概率或者統計概率的事物．粗略地来解释一下这句話的意思，可以这样說：我們所說的概率不是关于这些判断，而是关于一个理想实驗的可能結果．在我們談到概率之前，大家必須先承認一个特殊理想实驗的想象模型，譬如說扔錢币；观察扩散中一个質点；記录电話呼喚的次数．一开始就須承認什么是这个理想实驗所有的可能結果(即"样本空間")以及它們的概率．这个程序可以在力学里找到类似情况：在力学里引进了包含着2个，3个或17个質点的臆想模型，其中質点并沒有任何特性．同样，如果我們来分析扔錢币的实驗，我們理論的对象只是象"正面，正面，背面，正面，……"这些記号的序列，而不关心实际实驗的偶然情况．在这个体系里，象什么"明天出太阳的概率是多少"是不在考察之列的．假使我們要討論这个概率，我們势必先来确定一个实驗的(理想)模型，而这个实驗不免要說成"从无穷多个世界中任意地选取一个…"．造这样一个模型其实也不需要多大的想象力，但是

这样做是无聊而且毫无意义的.

天文学家談論太阳中心温度的測量，和到天狼星上的旅行. 这些似乎是不能办到的,但是作这种思考并不是沒有意义的. 同理,我們也不去忧虑理想的实驗能否实行;我們将要从事于分析抽象的模型. 在我們的思想深处,保持着概率的一个直观解释,这种解释在某些应用中获得实施的意义. 我們想象把实驗重复做很多次. 概率为 0.6 的事件終久可以期望在 100 次里出現 60 次. 这句話是有意地含糊其詞的，但是，对于較为初等的应用来說,它提供了一个足够形象的直观背景. 随着理論更精深的发展,实用意义和直观图象就变得更具体了.

4. 摘　　要

我們将要考虑一些理論模型,在这些模型中,概率象力学中的質量一样,是作为一些自由参数而出現的. 可以用各种各样不同的方法来应用它們，应用的技巧和直观能力随着理論的发展而发展.

这是其它数学学科中所采用的富有成效的标准手法. 沒有其它的手法可以满足正在发展着的概率論及其应用的各个分支的各种各样的需要.

使我們惋惜的是,直观的概率不能充分满足科学的需要,但是这是一个历史事实. 在第一章例(6.b)中，我們将要討論多个質点在多个盒里随机分布的問題. 适当的或者"自然"的概率分布对每个人来說似乎是完全清楚的，并且可以被物理学家毫无犹豫地接受. 然而,事情是这样,物理質点是沒有經过人类常識的訓練,而"自然"(或波茨曼(Boltzmann))分布有时必須代以爱因斯坦-波司(Einstein-Bose)分布，有时又必須代以弗米-迪拉克(Fermi-Dirac)分布. 直观論証不能說明为什么光子不同于質子，为什么它們不遵从一个"預先給定"的規律. 如果現在能找到一种論証,那也只能說明直观概念和理論一起发展罢了. 无論如何,即使对应用来說，自由和灵活都是重要的，而且把理論局限在一个固定的范畴

里是不利的.

有人叫嚷現代概率論是如此抽象和如此普遍化，以至难于应用. 这是只講实用的人們当初曾經反对馬克司威尔場論的挑战口吻. 这个論点可以为下列事实所駁倒：抽象的随机过程理論提供了一些新的未曾想象过的应用；現代的起伏理論所提供的新的知識又一次地与直观想法相违背，并且使人們去修正对实际的态度. 然而，这种辯論是沒有什么用处的，在昨天还認为是不合实际的东西今天就是实际的东西了，而明天将要成为理論的又被今天的实用的人們認为是毫无价值的游戏.

5. 历 史 小 記

对概率的統計或經驗观点主要是由馮·米賽斯(R.Von Mises)和費歇(R. A. Fisher)发展的. 样本空間的概念是由馮·米賽斯[2]引进的. 这个概念使得有可能把概率的严格的数学理論建立在測度論上. 在本世紀二十年代中，在許多作家的影响下，概率論的測度論方法逐漸形成. 現代概率的公理化处理，是由柯尔莫格洛夫(A. H. Колмогоров)[3]給出的. 我們将沿着这一綫索进行討論，但是，由于这一卷只討論离散概率的簡单情形，公理化系統这个名詞就显得太严謹了.

第一章　样本空間

1. 經驗背景

概率論的数学理論，从它和許多实际的和理想的实驗的关系上，获得了实用的价值，也获得了直观的意义．实际的和理想的实驗，例如有：扔一次錢币(鎳币)；扔一百次錢币；擲三顆骰子；理一付紙牌；用两付紙牌对点；玩輪盘賭；观察放射性原子的寿命或观察人的寿命；选取人为随机样本而观察其中慣用左手的人数；交配两种作物而观察它們后代的遺传型，或者概率論的数学理論，从它和一些現象的关系上，获得了实用的价值，也获得了直观的意义．一些現象，例如有：初生孩的性别；电話站中被占用的通話綫的数目；电話的呼喚次数；在电訊系統里面的随机噪声；生产过程的例行質量控制；意外事故的頻率；天空某一区域內双星的个数；在扩散过程中一个質点的位置．上列各項描述是含糊一点，要使概率論有意义，我們必須对所探討的实驗或观察的可能結果究竟是指什么而言取得一致的意見．

扔錢币掉下时不一定是正面或反面；它可能是滚掉了，也可能是笔直地站着．但是我們只承認正面和背面是扔錢以后仅有的可能結果．这样一来，理論要簡洁得多了，同时也不影响其应用．这种类型的理想化是标准的办法．測定一个原子的寿命或一个人的寿命而沒有誤差是不可能的，但是为了理論上的目的，我們不妨設想寿命是实实在在的一个数．問題发生了：确实能代表一个人的寿命的是些什么数值呢？有沒有生命不可逾越的最大年齡？是否一切年齡都是可以設想的呢？一方面，誰也不認为人能活到一千岁；另一方面，現时通行的保险业务对于人的可能寿命却不加任何上限．按照某些公式（死亡率表的現代制作法就是以这些公式为

基础的)算出来,千年不死的人在全人类中大約只占 $10^{10^{36}}$ 分之一，$10^{10^{36}}$ 这个数共含有 10^{27} 兆个零. 这句話，从生物学或社会学的角度看来,固然是毫无意义的,但是单純从統計上着眼，它和經驗当然沒有什么矛盾. 因为一世紀內出生的人数还不到 10^{10}. 要想用統計来考驗上述說法,就需要 $10^{10^{35}}$ 个世紀以上的一段时間，而这段时間比地球的寿命的 $10^{10^{34}}$ 倍还要大得多. 毫无疑問,这样小的概率和我們心目中的"不可能性"是沒有什么矛盾的. 你也許訊为,这种小概率的使用本身就是荒謬絕伦的. 其实不然,使用这种小概率非但沒有坏处,而且还可以簡化公式. 再說,如果我們眞的把活一千年的可能性排除掉，我們势必承訊一个最大年齡限 x 存在,說人能活 x 年而不能活 x 年零两秒,这种說法决不会比无限寿命的說法更能講得通些.

任何理論都必然含有理想化,对于我們来說,第一个理想化是关于实驗或观察的可能結果. 如果我們要为实驗制作一个抽象模型,我們必須一开始就作出决定: 这实驗的可能結果是由哪些东西构成的.

为了統一术語起見，我們把实驗或观察的結果叫做**事件**. 这样一来,我們就可以談論"扔 5 个錢币至少出現 3 个正面"的事件. 同样,分桥牌[1]的"实驗"，其結果可以是"北家拿到 2 张爱司(ace)"的"事件". 一个样本的成分 (例如"85 人組成的样本中 2 个慣用左手") 和一个測量的結果(例如"温度 120°", "7 条电話綫占用") 也都叫作事件.

我們要区分复合(或可分解的)事件和簡单(或不可分解的)事件. 例如, 要是擲二个骰子使"总和为 6",那就是使成为"(1, 5)或(2, 4)或(3, 3) 或 (4, 2) 或(5, 1)"，而这个事例把 "总和为 6" 的

---

1) **桥牌和玩扑克的定义**: 一付桥牌共 52 张,分 4 种花色,每种花色有 13 张. 每一种花色有 13 个不同的面值: 2, 3, …, 10, "賈克"(jack),"坤"(queen),"老开"(king),"爱司"(ace). 4 种花色称为黑桃、梅花、紅心、方块, 前两种是黑色的而后两种是紅色的. 同面値的牌称为同点. 所謂玩桥牌,就是把整付牌分发給 4 家,这四家称为"东"、"南"、"西"、"北",每家各得 13 张. 至于玩扑克的意义是: 从一付牌里每家各取 5 张,进行組合.

事件分解成五个簡单事件. 同样,“两个奇数点”事件就分解为“(1,1)或(1,3)或……或(5,5)”九个簡单事件. 注意: 如果擲出的結果是(3,3),那么,这个結果既包含在事件“和为 6 点”又包含在事件“二个奇数点”之內;而这两个事件不是互斥的,它們可以同时发生. 作为第二个例子,我們来考虑人的寿命. 每一个特殊数值 x 代表一个簡单事件,“此人 50 多岁”代表 x 在 50 到 60 这一事件. 用这种办法,每一个复合事件都可以分解为一些簡单事件,也就是說,每一个复合事件是一些簡单事件的集合.

如果要在理論上很明确地討論“实驗”或者“观察”,必須首先約定: 簡单事件代表可以想象的結果,它們确定理想的“实驗”. 习惯上,这些簡单事件叫作**样本点**,或干脆就叫“**点**”. 由定义得知:(理想)实驗的每一个不可分解的結果可用一个且只能用一个样本点来表示. 所有这些样本点的全体称之为**样本空間**. 于是,牽涉到給定的(理想)实驗的一切事件,都可以用样本点来表达.

在把这些基本的約定形式化以前,我們討論几个以后常要用到的例子.

2. 实 例

(a) 三个球在三个盒中的分布. 表 1 列出了三个球放入三个盒中的全部可能結果.

表 1

1. $\{abc \mid - \mid -\}$	10. $\{a \mid bc \mid -\}$	19. $\{- \mid a \mid bc\}$
2. $\{- \mid abc \mid -\}$	11. $\{b \mid a\,c \mid -\}$	20. $\{- \mid b \mid a\,c\}$
3. $\{- \mid - \mid abc\}$	12. $\{c \mid ab \mid -\}$	21. $\{- \mid c \mid ab\}$
4. $\{ab \mid c \mid -\}$	13. $\{a \mid - \mid bc\}$	22. $\{a \mid b \mid c\}$
5. $\{a\,c \mid b \mid -\}$	14. $\{b \mid - \mid a\,c\}$	23. $\{a \mid c \mid b\}$
6. $\{bc \mid a \mid -\}$	15. $\{c \mid - \mid ab\}$	24. $\{b \mid a \mid c\}$
7. $\{ab \mid - \mid c\}$	16. $\{- \mid ab \mid c\}$	25. $\{b \mid c \mid a\}$
8. $\{a\,c \mid - \mid b\}$	17. $\{- \mid a\,c \mid b\}$	26. $\{c \mid a \mid b\}$
9. $\{bc \mid - \mid a\}$	18. $\{- \mid bc \mid a\}$	27. $\{c \mid b \mid a\}$.

其中每一个排列都代表一个簡单事件,即是一个样本点. 事

件 A："在一个盒内放一个球以上"为第 1 到第 21 个排列的总体，我們說事件 A 是由第 1 到第 21 个样本点所构成的集合．类似地，事件 B："第一个盒是不空的"是第 1，第 4 到第 15，第 22 到第 27 这几个样本点构成的集合．事件 C："A 和 B 都发生"是由第 1，第 4 到第 15 共 13 个样本点构成的集合．在这个特例中，27 个样本点中的每一个都或者属于 A 或者属于 B(或者旣属于 A 又属于 B)；因此，事件"或者 A 或者 B 或者二者都发生"就是样本空間，因此它必然发生．事件 D："A 不发生"是由第 22 到第 27 个这六个样本点所构成，它可以用下述条件来描述：沒有一个盒是空的．事件："第一个盒是空的而其它的盒沒有放一个球以上"是不可能发生的，因为沒有一个样本点能滿足这些条件．

(b) r 个球在 n 个盒中的分布． 对于 r 个球分布在 n 个盒中的一般情形完全可以用类似的办法来进行研究，只不过这时的排列的个数随 r 和 n 的增加而增加．当 $n=3$，$r=4$ 时，样本空間由 81 个样本点构成，当 $r=n=10$ 时，样本空間共有 10^{10} 个样本点；而造一个完整的表就要有約十万卷篇幅．

我們用上面这个例子来說明一个重要的事实：即是样本点的性質和我們的理論是无关的．对于我們来說，样本空間(与在其中定义的概率分布一起)决定了理想的实驗．我們应用球和盒这种形象的語言，但是，同一个样本空間可以允許有很多种不同的实际解释．为了說清这一点并为了今后的应用，我們在这里抄录一些其直观背景很不相同的实驗，然而，抽象地看，它們都等价于 r 个球分布于 n 个盒中的模型．在这些情形中，其合理的賦概是不完全一样的，后面我們将要重新討論．

(b, 1) 生日． r 个人的生日的可能情形相当于 r 个球放入 365 个盒中的不同排列(假定一年有 365 天)．

(b, 2) 不幸事件． 如果把 r 个意外事件按其发生在星期几来分类的話，則它等于 r 个球放入 $n=7$ 个盒中．

(b, 3) 打 n 个靶．子弹相当于球，靶相当于盒．

(b, 4) 抽样． 把 r 个人按其年龄和职业来分类． 于是类就

相当于盒而人就相当于球.

(b, 5) 生物学中的照射. 当光綫射到眼网膜中的細胞时,光质点相当于球,而实际的細胞就是我們模型中的"盒". 类似地,在研究照射的遺传效果时,染色体相当于盒,而 α 质点相当于球.

(b, 6) 在宇宙射綫的实驗中,击中蓋格(Geiger)計数器的质点相当于球,而計数函数相当于盒.

(b, 7) 一部电梯, 开始有 r 个乘客, 它在 n 层楼中每一层都停. 乘客走出电梯的各种不同的排列与 r 个球放入 n 个盒中的各种不同排列相同.

(b, 8) 骰子. 擲 r 个骰子的可能結果相当于把 r 个球放入 $n=6$ 个盒中. 如果是扔錢币,則对应的盒只有 $n=2$ 个.

(b, 9) 随机数. r 个数字所构成的序列的各种可能的次序相当于 r 个球(=位置)放入 10 个称为 0, 1, $\cdots$, 9 的盒中的可能分布.

(b, 10) r 个人的性别分布. 这时我們有 $n=2$ 个盒以及 r 个球.

(b, 11) 贈劵 (coupon)* 的收集. 不同种类的贈劵相当于盒; 收集的贈劵代表球.

(b, 12) 桥牌中的爱司. 四个玩牌者代表四个盒, 我們有 $r=4$ 个球.

(b, 13) 遺传因子的分布. 每一个生物(人, 植物或动物)的后代都从其祖先那里継承一些遺传因子. 如果一个特殊的遺传因子可以有 n 种不同的方式 $A_1, \cdots, A_n$,則后代可以按其遺传因子的类型来分类. 后代相当于球, 遺传因子的类型 $A_1, \cdots, A_n$ 相当于盒.

(b, 14) 化学. 假定长鏈聚合物与氧反应. 每一个鏈都可能和 0, 1, 2, ……个氧分子起反应. 这里参加反应的氧分子相当于球,而聚合物的鏈相当于盒.

---

* 在資本主义国家里,为了推銷商品追求利潤,在出售商品中附以贈劵, 規定在收集了某些花式或号碼后可免費获得贈品——中譯本編者.

(b, 15) 显影液的理論. 在一个照象底板上涂上一层容易感光的液体，当这种液体的粒子被 r 个光子击中时，它就起反应. 为了区分黑白，我們必須知道多少个粒子(想象为盒)被 r 个光子所击中. 这里，我們得到一个占位問題，粒子相当于盒，而光子相当于球. (当然，实际問題是很复杂的，因为底板上的液体的粒子的感光性強弱是不一样的.)

(b, 16) 印錯. r 个錯誤在一本具有 n 頁的书中的一切可能的分布相当于 r 个球放入 n 个盒中的一切可能分布，不过 r 必須小于每一頁的字数.

(c) 球为不可辨的情形. 讓我們回到例(a)，并且假定那三个球是不可辨别的. 这意味着：正如 4, 5, 6 中的三种不同的排列都分不清了，因此表 1 变为表 2.

表 2

1. {*** | — | —}	6. { * | ** | —}
2. {— | *** | —}	7. { * | — | ** }
3. {— | — | ***}	8. {— | ** | * }
4. {** | * | —}	9. {— | * | ** }
5. {** | — | * }	10. { * | * | * }.

表 2 确定了下述理想实驗的样本空間：把三个不可辨别的球放入三个盒中，而且同样我們可以采用 r 个球放入 n 个盒中.

球在实际上是否可辨别与我們的理論不相干. 甚至当它們可以辨别时，我們也可以作不可辨别的来处理. 桥牌中的爱司(例(b, 12))或者电梯中的人(例(b, 7))都是可以辨别的，但是經常把它們当作不可辨别的来处理反而很方便. 例(b, 8)中的骰子可以涂上顏色使之可辨别，但是，当我們討論一些特殊的問題时，应用可辨别的或不可辨别的球的模型可以根据我們的特定的目的和方便来决定. 問題的性質决定了我們如何选择，不过，在任何情形，只有当适当的模型选定以后，也就是說，当样本空間选定了以后，我們的理論才能开始.

在上面的模型中，我們考虑不可辨别的球，不过表 2 仍然是指

第 1,第 2,第 3 个盒,而且它們的次序还是不可缺少的. 我們可以进一步假定盒也是不可辨别的(例如,盒可以随机地选取而不考慮其內容). 当球和盒都是不可辨别的时候,所有可能的排列只有三种,即是{***|—|—}, {**|*|—}, {*|*|*}.

(d) 抽样. 假定为了估計吸烟的人数,我們任意选取 100 个人作为样本. 在这里,我們只对其中吸烟的人数 x 感兴趣,而 x 可以是 0 到 100 之間的任意一个数. 这样一来, 我們可以确定这个样本空間是由 $x=0, 1, 2, \cdots, 100$ 等 101 个 "点" 所組成的. 每个个别的样本或观察的結果完全可以用对应的点 x 来代表它. 現在举一个复合事件的例子: "样本中的人多半数吸烟". 这个事件意味着,实驗的結果表明, 在 $x=51, 52, \cdots, 100$ 等 50 个事件中有一个发生; 究竟发生的是那一个呢? 这里并沒有交待. 同样,样本的每个性質,都可以通过列举它所对应的情况或样本点来表示. 为了术語上的統一, 我們宁愿說事件而不說样本的性質. 用数学的术語来說: 一个事件就是与之对应的所有的样本点的集合.

(e) 抽样(續). 現在我們对这 100 个人,不但要按吸烟与否来分类,而且还要分別他們的男女性别. 于是,現在的样本点須要由男女吸烟者,男女非吸烟者的人数来表示,也就是由整数的四元組(M_s, F_s, M_n, F_n)来描述. 我們可以把这样的四元組的全体作为样本空間: 其中的每个整数都在 0 与 100 之間, 而每組四个数的和为 100. 所謂在一个样本中,"男子中吸烟的人相对地多于女子中吸烟的人",意思就是說: 在这个样本中,比值 $\frac{M_s}{M_n}$ 大于 $\frac{F_s}{F_n}$. 譬如象"点"(73, 2, 8, 17)就具有这个性質,而"点"(0, 1, 50, 49)則不然. 在原則上,任何事件都可以把具有所須特性的全部四元組列举出来作为描述.

(f) 扔錢币. 如果把一个錢币扔 3 次,这时, 样本空間就由 8 个点构成. 如果以 H 表示正面, T 表示背面,那末, 这 8 个点就可以用 HHH, HHT, HTH, THH, HTT, THT, TTH, TTT 来

代表."其中至少有2次出現正面"的事件 A, 可以由这8个点中前4个所成的集合来表示."恰巧有1次出現背面"的事件 B, 是指 HHT 或 HTH 或 THH, 所以我們就說 B 包含着这3个点.

(g) **配偶的年龄.** 保险公司对于配偶年龄的分布有兴趣.設我們以 x 代表丈夫的年龄, y 代表妻子的年龄. 于是每作一次考察就可以得到一个数对 (x, y). 对应于考察的整个样本空間,就可以用 xy 平面的第一象限来表示. 于是每一个 $x>0$, $y>0$ 的点都是这个样本空間中的一个样本点."丈夫的年龄大于40岁"的事件 A,可以用直綫 $x=40$ 右边的全部点来代表;"丈夫的年龄比妻子大"的事件 B, 可以用 x 軸,与第一象限的分角綫 $y=x$ 之間所夹的角状区域来代表,也就是說可以用 $x>y$ 的点的全体来代表;"妻子的年龄大于40岁"的事件 C, 可以用第一象限中位于直綫 $y=40$ 以上的部分来代表. 至于二对配偶年龄的联合分布的几何表示,就需要用到四維空間了.

(h) **相空間.** 在統計力学里,体系的每个可能"状态"被称为相空間中的一个"点". 这仅仅只是术語上的不同,其实,相空間就是样本空間,它的点也就是样本点.

3. 样本空間.事件

根据前面所述,可以清楚地看出, 我們不討論概率則已,如果要提到概率,我們总是把它結合着所給定的样本空間(或者用物理上的話来說: 結合着某个理想的实驗)来談. 我們由样本空間和它的点的概念出发,今后我們認为,它們是已經給定了. 它們是概率論中的原始的无定义的概念, 正如"点"和"直綫"是欧氏几何的公理化处理中的无定义的概念一样. 样本点本身的內在性質与我們的理論毫无关系. 样本空間在如下意义下提供了一个理想实驗的模型,按定义来說就是: 实驗中每个可能設想的結果,都完全可以由唯一的样本点来描繪. 只有从实驗的每个結果都能清楚地判定某一事件 A 是发生或者不发生时, 談起这个事件 A 来才有意义.代表"A 发生了"这个事件結果的所有那些样本点的集合, 就完全

地描繪着"A 发生了"的这个事件. 反过来,包含着一个或多个样本点的任意給定的集合 A 都可以当作一个事件来加以討論；这个事件是否发生，就由实驗的結果是否由 A 中的点所表示而定. 所以事件这个詞的定义,就完全等同于样本点的一个集合. 我們說一个事件 A 包含着某些点(或由某些点构成)，就是說这些点表示在理想实驗中能使 A 发生的那些結果.

例. 在例(2.a)的样本空間中，考慮由点 1，7，13 所构成的事件 U. 这是一种形式的、直接的定义,但是,U 可以用很多等价的方法来描述. 例如,U 可以定义为下面三个条件都滿足的事件：(1)第二个盒是空的；(2)球 a 在第一个盒中;(3)球 b 不在 c 后出現. 这三个条件的每一个都确定一个事件. 由条件(1)所定义的事件 U_1 由点 1，3，7—9，13—15 所构成. 由条件(2)所定义的事件由点 1，4，5，7，8，10，13，22，23 所构成,而由条件(3)所定义的事件由点 1—4，6，7，9—11，13，14，16，18—20，22，24，25 所构成. 事件 U 也可以考虑作三个事件 U_1, U_2, U_3 同时发生.

术語"事件"和"样本点"有着直观上的輔助作用，而这些概念等价于在各数学分支中常見的点集和点的概念.

在前面的例子和例子(2.a)中，我們看到了，由两个或者两个以上的事件可以定义一个新的事件. 由于这些例子的启发，我們引进事件的形式代数的概念(即是点集的代数的概念).

4. 事件之間的关系

現在,我們假定一个任意然而又是固定的样本空間 $\mathfrak{S}$ 已經給定.

定义 1. 我們用符号 $A=0$ 来表示事件 A 不包含任何样本点(即是不可能事件). 0 必須了解为一个符号而不是数.

相对于任何一个事件 A，都有以"A 不发生"为条件来定义的另一事件,它包含着所有不属于 A 的点.

定义 2. 样本空間中一切不属于事件 A 的点所构成的事件称为 A 的补(或称 A 的非)事件,并以 A' 記之. 特別地，$\mathfrak{S}'=0$.

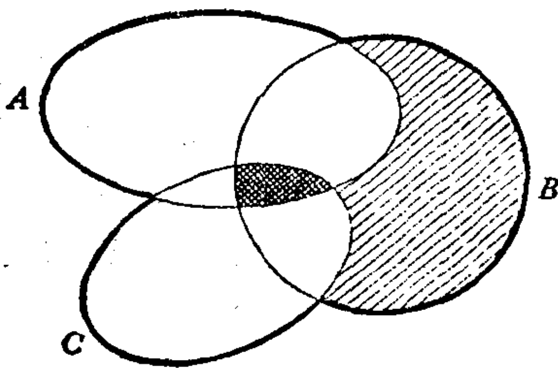


图 1

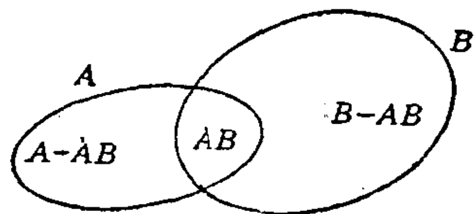


图 2

图 1 和图 2 說明事件之間的关系．在图 1 中以粗綫包圍的区域是幷集[1] $A\cup B\cup C$．中間黑色的三角形区域是交集 ABC．有浅色阴影具有月亮形的区域是 B 与 $A\cup C$ 的补集之交．

对于任意两个事件 A 及 B，总可以伴随着两个新的事件，它們是以条件"A 和 B 同时发生"和"A 或 B 发生，或两者同时发生"来定义的．这两个事件分别以 AB 和 $A\cup B$ 記之．事件 AB 包含 A 和 B 中所有的公共的样本点．如果 A 和 B 互相排斥，則就沒有公共点，因而事件 AB 成为不可能事件．这时，我們就用方程

$$AB = 0 \tag{4.1}$$

表示，幷讀作"A 和 B 互不相容"．事件 AB' 表示 A 和 B' 同时发生，换句話說，就是"A 发生而 B 不发生"．同样，$A'B'$ 表示 A 和 B 都不发生．事件 $A\cup B$ 表示 A 和 B 中至少有一个发生，它包含既不属于 A 又不属 B 的点以外的全部样本点．

在概率論里，事件 AB 是用 A 和 B 同时发生的話来描述的．如果要用标准的数学术語来說，那末 AB 就称为 A 和 B 的(邏輯上的)交集．同样，$A\cup B$ 就称为 A 与 B 之幷．这些概念可以推广到好几个事件 $A, B, C, D, \cdots$ 的情况．

定义 3. 对每一組事件 $A, B, C, \cdots$，可以把二个新事件定义如下：第一个事件是由属于全部給定的集合中的样本点的全体所組成的，我們用 $ABC\cdots$ 来表示它，幷且叫它为 $A, B, C, \cdots$ 之

1) 幷集，也可以称为和集——譯者注．

交[1](同时实現). 第二个事件是由至少屬于給定的各集合中之一的样本点的全体所組成,我們用 $A \cup B \cup C \cup \cdots$ 来表示,并称之为給定的各集合之并(至少有一个实現). 在事件組 $A, B, C \cdots$ 中,如果任何二个都沒有公共点,卽 $AB=0$, $AC=0$, $\cdots$, $BC=0$, $\cdots$, 則称这組事件是两两互斥的.

我們还需要用一个記号来表达这样的命題:"当 B 不发生时, A 不可能发生";或者說"A 的发生蘊含着 B 的发生". 这意思是說: A 的每个点都包含在 B 之中. 不妨作一个直观的想象,譬如所有的母亲所成的集合是全体妇女的一部分:所有的母亲都是妇女,但并不是所有的妇女都是母亲.

定义 4. 記号 $A \subset B$ 及 $B \supset A$ 是等价的,这表示 A 的每一个点,都包含在 B 之中,我們分别讀为"A 蘊含 B"及"B 被 A 蘊含". 如果上述蘊含关系成立,我們就用 $B-A$ 来代替 BA',它表示事件"B 发生而 A 不发生".

事件 $B-A$ 包含着所有在 B 中而不在 A 中的那些点. 根据这个記号,我們可以写出 $A' = \mathfrak{S} - A$ 和 $A - A = 0$.

例. (a) 若 A 和 B 互不相容,則 A 的发生蘊含着 B 不发生,反之亦然. 因此, $AB=0$ 也就是 $A \subset B'$ 或 $B \subset A'$.

(b) 事件 $A-AB$ 表示 A 发生而并不是 A 和 B 同时发生,因此 $A-AB=AB'$.

(c) 在例 (2.g) 中,事件 AB 意味着丈夫的年龄大于 40 岁而且还比他的妻子年龄大; AB' 意味着丈夫的年龄大于 40 岁而不比他妻子大. AB 可以用由 x 軸, $x=40$ 和 $y=x$ 所构成的无穷的梯形区域来表示,事件 AB' 可以用 $x=40$ 和 $y=x$ 所构成的三角形区域来表示,后者 $y=x$ 包含在区域內. 事件 AC 表示丈夫和妻子的年龄都大于 40 岁,事件 $A \cup C$ 表示夫妻之中至少有一个大于 40 岁, $A \cup B$ 表示或則丈夫大于 40 岁,如果不然,至少

---

1) 两个或两个以上的事件的交的标准符号是 $A \cap B$ 或者 $A \cap B \cap C$ 等等. 为了某些特殊的目的,这种符号更好一些,因此,在第二卷里我們就采用这种符号. 不过,为了排印的簡便起見,現在我們仍用 AB 或 ABC 等符号.

比妻子大(用一般的語言来說,丈夫的年龄超过 min (40,妻子的年龄)).

(d) 在例(2.a)中,令 E_i 表示第 i 个盒是空的事件 ($i = 1, 2, 3$). 类似地,令 S_i, D_i, T_i 分别表示第 i 个盒放一个,两个,三个球. 于是 $E_1E_2 = T_3$, $S_1S_2 \subset S_3$, $D_1D_2 = 0$. 注意 $T_1 \supset E_2$, 等等. 事件 $D_1 \cup D_2 \cup D_3$ 是由条件"至少有一个盒放两个球"所确定的.

(e) 桥牌(参看第一节脚注). 令 A, B, C, D 分别表示北家至少有一张爱司,南家至少有一张爱司,东家至少有一张爱司, 西家至少有一张爱司的事件. 显然,至少有一个玩牌者拿到一个爱司,所以这四个事件至少有一个事件一定发生,因此$A \cup B \cup C \cup D = \mathfrak{S}$ 是全样本空間. 事件 $ABCD$ 当且仅当每一个玩牌者都拿到一个爱司时才发生. 事件"西家有四个爱司"的发生意味着事件 A, B, C 都不发生;即是 A', B', C' 都发生,或者 $A'B'C'$ 发生.

(f) 在例子(2.g)中,我們有 $BC \subset A$;即是說,"如果丈夫的年龄大于妻子(B)而且妻子的年龄大于 40 岁,則丈夫的年龄大于 40 岁(A)". 事件 $A - BC$ 如何用語言来描述?

5. 离散样本空間

最簡单的样本空間, 莫过于那些只含有限多个(譬如說 n 个)点的空間.如果 n 很小(例如在扔几个錢币的場合),則(样本)空間就很容易具体化. 但在玩桥牌的游戏中,要具体化牌的分配情况所构成的(样本)空間就比較复杂些. 然而无論如何,我們还是可以这样来設想;把每个样本点描写在一个筹碼上,于是这些筹碼的全体,就代表了这个样本空間. 事件A(例如"北家有 2 个爱司")就可以用筹碼中某一部分来代表; 而 $\bar{A}$ 可以用剩下来的那些筹碼来代表. 只有从这儿再前进一步, 我們才能設想到具有无穷多个筹碼的情况,或者具有无穷点 $E_1, E_2, E_3, \cdots$的一个样本空間.

例. (a) 設我們来扔一个錢币, 一直扔到它出現正面为止. 于是,这个样本空間的点就是(以H表正面,T表背面,以下常用这二个記号,不再注出): $E_1 = H$, $E_2 = TH$, $E_3 = TTH$, $E_4 = TTTH$,

等等. 至于H永不发生的可能性可以有两种不同的考虑方式，或者認为它是可以設想的,或者認为那是不可設想的. 如果我們認为那是可以設想的,那末,这个可能性就用点E_0来代表.

(b) 假定有三个人 a, b, c 在作游戏,比如說下棋,其規則如下: 开始时a和b对下,而c在局外. 誰輸了誰就被c来代替,因而在第二次游戏中赢家和c对下,而輸家在局外. 这个游戏用这种方式一直作到有一个人接連赢两次时为止. 为了簡单起見,每一次游戏中不分胜負的可能性不加考慮. 我們的游戏的可能結果可以由下面的方案来表示:

$$(*)\quad \begin{matrix} aa,\ acc,\ acbb,\ acbaa,\ acbacc,\ acbacbb,\ acbacbaa,\ \cdots \\ bb,\ bcc,\ bcaa,\ bcabb,\ bcabcc,\ bcabcaa,\ bcabcabb,\ \cdots. \end{matrix}$$

此外,可以假想沒有一个人接連赢两次,这意味着游戏按下述方式之一无穷地进行下去:

$$(**)\quad acbacbacbacb\cdots,\quad bcabcabcabca\cdots.$$

我們这个理想的"試驗"的样本空間由(*)和(**)所确定，而且是无穷的. 显然,其样本点可以排成一个簡单的序列,在(**)中的两个点排在最前面,接着把(*)中的点子按$aa, bb, acc, bcc, \cdots$的次序継續排列下去.（这个例子将要在問題 5 和 6, 第五章例子(2.a);第十五章問題 5 中継續討論.）

定义. 如果一个样本空間只包含有限多个点,或者虽有无穷多个点,而这些点可以排成一个簡单的序列$E_1, E_2, \cdots$的話,我們就說这个样本空間是离散的.

并不是每一个样本空間都是离散的. 根据熟知的定理(康托(G. Cantor) 定理),由一切正数所組成的样本空間就不是离散的. 这里,让我們把力学上大家所熟悉的特性对比着来叙述: 在力学上,首先考虑的往往是离散的质点,而每个个别的点都带着有限的质量. 这个概念与质量連續分布的概念有着显著的区别，因为后者是說: 每个个别的点都是沒有质量的. 在第一种情况下,要求体系的总质量，只要把个别点的质量加起来，很容易的就可以得到;而在第二种情况下,总质量需要按质量的分布密度进行积分来

計算．情形十分相似，在离散的样本空間中，也只要用加法就可以算出事件的概率，而在其他样本空間中，也要用到积分．对于这二种情况，除了需要不同工具而外，本質上并沒有什么区别．为了从反映現实的概率来考虑，而不为技巧上的困难所限制，我們只討論离散的样本空間．以后会看到，即使只討論这种特殊的情况，也会得出很多有趣味而又重要的結果．

在这本书里，我們只考虑离散的样本空間．

6. 离散样本空間中的概率：預备

各种事件的概率都是一些数，这些数的性質犹如几何学中的距离与力学中的質量．概率論假定概率的数值是給定的，但是关于它們的实在数值等于什么，怎样測量出来的，却用不着作任何的假定．一些重要的应用具有定性的性質，并沒有用到数值．概率論的一般結論应用到各个方面，犹如几何定理作为物理理論及工程应用的基础一样．当眞正需要知道概率的数值时（这种情况并不常見），那末所用的方法和測量距离一样是变化多端的（在測量距离时，木匠师傅，測量員，飞行員和天文工作者所用的方法很少有相似之处）．例如，在我們的书文里，我們要考虑扩散常数，它是概率論中的一个概念．要想得到扩散常数的数值，那就需要它和其它理論联系起来的物理思考；而直接的測定是不可能的．相反，为制作死亡率表而需要的概率数值，却是得自比較原始的观測資料的．在很多应用方面，为了确定概率，或者把理論与观測作比較，往往要用到一些巧妙的統計方法，而这些方法却又建筑在概率論的精深理論上．換句話說，概率的直观意义是清楚了，但是只当理論向前进展了，才能够看到它是怎样被应用．光靠着給概率下这个或那个“定义”，对于实践来說是远远不够的．

扔一个“均匀的”錢币时，我們毫不犹疑地認为正面和背面的概率都是 1/2．这也就是說，如果扔錢 n 次，那末 2^n 个可能結果都有相同的概率．从理論的观点来看，这无非是一个約定．許多人認为这个約定是不可避免的，而且是唯一可能的一个．另一方面，曾

經有一批哲学家和統計学家不滿足于这个約定，主张从相矛盾的假設出发(自然界的齐一性或不齐一性). 也有人主张，1/2 这个概率是从經驗得来的. 然而，每当使用精深的統計方法来检驗实际扔錢的結果的时候，結論总归是：正面和背面的出現不是等同概然的. 尽管"絕对均匀的"錢币幷不存在，我們还是坚持有一个"理想"的錢币. 这不单单是为了它的邏輯簡单性，主要是因为它是有用的而且是能用的. 一則，在許多应用方面，这个模型已經把現实描写得够准确的了. 更重要的却是这样一个經驗事实，即是：每当与我們的模型偏离时，总是伴随着，例如，重心的位置不正的現象. 这样一来，就算理想模型永远不会絕对准确地实現，它倒反而是非常有用的了，因为对它的偏离可以提醒我們去发掘潛伏着的不正常現象. 举例来說，在現代統計的質量控制法[創始人是休哈特(Shewhart)]里面，我們就使用理想化的概率模型，以便在对于模型有显著偏离的时候，就去查究"可推源的原因"，从而及早地清除掉卽将来临的机器故障和生产过程中的不正常現象.

类似的討論可以应用到其它的場合. 例如，分配桥牌方式的可能的数目大約在 10^{30} 左右. 通常，我們約定把它們作为等同概然的. 要想对这个約定作一次核对，那么必須进行 10^{30} 次以上的試驗——相当于一个人，日夜不停地每秒鈡玩一局，大約要玩一百亿亿年(10^{14})的局数. 虽然如此，这个假定的后果还是可以用实驗来驗証的，譬如說，用观察手里拿到 2 个"爱司"的頻率作实驗，就可以来驗証它. 观察的結果是：只需把牌洗得特别匀，理想化的模型确是将經驗描写得大致不差. 更重要的是：当理想化的模型发生問題时，那就有可能使我們进而去发現矛盾存在的"可以推源的原因"，例如，需要改善洗牌的方式. 这些都是不怎么重要的例子，但它們还是显示了假設模型的用途. 更有趣味的情形，在以后还会談到.

例. (a) 可辨别的球. 在例(2.a)中，自然地假定所有的样本点都是等可能的，卽是每一个样本点都具有概率 1/27. 我們可以从这个定义出发来研究其結果. 我們的模型是否和实际实驗很

相合将要依赖所考虑的現象的类型.在某些应用中,从物理意义出发,可以作"等概性"的假設;但是,在另外一些应用中,它之被引进只是作为一般情形的一个最簡单的模型,甚至有时很明显地看出它只是一个粗糙的第一步的近似,我們仍然作这种假設.(例如,考虑例子(2.b,1)生日;(2.b,7)电梯問題;或者(2.b,11)贈劵收集.)

(b) 不可辨别的球:波司-爱因斯坦統計. 現在,我們回过来考虑三个不可辨别的球分布于三个盒中的例(2.c). 球之不能辨别是不影响实际的物理試驗的;从物理上看,它們仍然有 27 种不同的可能,虽然其中只有 10 种可以辨别. 这种考虑使我們对表 2 中的样本点可以按照下述办法賦概:

点数:	1	2	3	4	5	6	7	8	9	10
概率:	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{1}{27}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{2}{9}$.

只有当这个結論在例(2.b)中所抄录的許多应用中是正确的时候,这种賦概法才是合理的. 在一个长时期內,我們的結論都被接受下来,沒有产生什么問題,而且它是作为統計力学中的 r 个球分布于 n 个盒中的馬克司威尔-波茨曼統計的推导的基础. 但是,当波司和爱因斯坦証明某些質点遵从波司-爱因斯坦統計时(詳細情形,参看第二章第 5 节),才引起人們的惊异. 在我們的情形下,$n=r=3$,波司-爱因斯坦給这十个样本点中的每一个都賦以概率 1/10.

这个例子将要説明,在相同的样本空間中可以有不同的賦概法,并且还将説明理論和实驗之間的错綜复杂性. 特别,它告訴我們不要过分地信賴前面的結論,而要准备接受新的非先驗的方案.

(c) 扔錢币. 等概性的假設常常需要眞正的实驗的記录来説明.在实践中,每一个錢币总不能絕对均匀,但是,可以設計一个物理实驗,使其与理想的扔錢币的模型非常近似. 为了給出一个期望的波动概念,我們給出一个对应于 10000 次扔錢币的假想的試驗[1]. 表 3 包含了每 100 次試驗出現正面的次数,而每 100 次試

1) 这个表眞实地記录了具有100,000正态离差的1,000,000个随机数中偶数的頻数.

表 3

試驗次数	正		面			次		数			总数
0—1000	54	46	53	55	46	54	41	48	51	53	501
—2000	48	46	40	53	49	49	48	54	53	45	485
—3000	43	52	58	51	51	50	52	50	53	49	509
—4000	58	60	54	55	50	48	47	57	52	55	536
—5000	48	51	51	49	44	52	50	46	53	41	485
—6000	49	50	45	52	52	48	47	47	47	51	488
—7000	45	47	41	51	49	59	50	55	53	50	500
—8000	53	52	46	52	44	51	48	51	46	54	497
—9000	45	47	46	52	47	48	59	57	45	48	494
—10000	47	41	51	48	59	51	52	55	39	41	484

驗对应于把一个錢币扔 100 次. 其总数为 4979. 看了这些数字以后,讀者很可能产生一些含糊的感覺: 这說明什么呢? 为了判断对什么范围来說能使經驗事实与抽象模型一致, 需要一些更高深的理論. (在第三章第 7 节我們将要回过来考虑这件事情.)

7. 基本定义和规則

基本的約定. 給定一个含有样本点 $E_1, E_2, \cdots$ 的离散样本空間 $\mathfrak{S}$,我們假定对每一个点 E_j 都賦以一个数,这个数称为 E_j 的概率,記以 $\mathbf{P}\{E_j\}$. 这些数都是非負的而且满足

$$\mathbf{P}\{E_1\} + \mathbf{P}\{E_2\} + \cdots = 1. \tag{7.1}$$

注意,我們并不排除一个点的概率为 0 的可能性,这种約定可能产生一些人为的但却是必需的复杂性. 在离散的样本空間中,零概率在实际中解释为不可能,因此,把已知其概率为 0 的样本点从样本空間中除去是无妨的. 然而,事前往往不知道其概率的数值,因此,判断一个样点的概率是否是正的需要深刻地推理.

定义. 任何一个事件 A 的概率 $\mathbf{P}(A)$ 是 A 中所包含的样本点的概率的总和.

基本方程(7.1)說明: 全样本空間的概率为 1,即是 $\mathbf{P}\{\mathfrak{S}\}=1$. 由此推出,对任何一个事件 A, 都有

$$0 \leqslant \mathbf{P}\{A\} \leqslant 1. \tag{7.2}$$

現在我們考慮两个任意的事件 A_1 和 A_2. 为了計算或則 A_1 发生,或則 A_2 发生,或則两个都发生的概率 $\mathbf{P}\{A_1 \cup A_2\}$, 我們把或者屬于 A_1, 或者屬于 A_2 的样本点的概率加起来,不过每一个样本点只算一次. 因此,我們有

$$\mathbf{P}\{A_1 \cup A_2\} \leqslant \mathbf{P}\{A_1\} + \mathbf{P}\{A_2\}. \tag{7.3}$$

如果 E 是一个旣屬于 A_1 又屬于 A_2 的样本点, 則 $\mathbf{P}\{E\}$ 在右边出現两次,而在左边只出現一次. 因此,右边比左边大 $\mathbf{P}\{A_1A_2\}$, 从而我們有下述簡单的但却是重要的

定理. 对于任何两个事件 A_1 和 A_2, 或 A_1 发生,或 A_2 发生,或 A_1, A_2 都发生的概率由

$$\mathbf{P}\{A_1 \cup A_2\} = \mathbf{P}\{A_1\} + \mathbf{P}\{A_2\} - \mathbf{P}\{A_1A_2\} \tag{7.4}$$

所給出. 如果 $A_1A_2 = 0$, 也就是說 A_1 和 A_2 是互斥的,則(7.4)化为

$$\mathbf{P}\{A_1 \cup A_2\} = \mathbf{P}\{A_1\} + \mathbf{P}\{A_2\}. \tag{7.5}$$

例. 把一个錢币扔两次. 我們把四个点 HH, HT, TH, TT 作为样本空間, 每一点賦以概率 1/4. 令 A_1 和 A_2 分別表示事件"第一次出現正面"和"第二次出現正面". 于是 A_1 包含了 HH 和 HT, A_2 包含了 TH 和 HH. 因此 $A_1 \cup A_2$ 包含了三个点 HH, HT 和 TH, 而 A_1A_2 只包含一个点 HH. 因此

$$\mathbf{P}\{A_1 \cup A_2\} = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}.$$

n 个事件中至少有一个发生的概率 $\mathbf{P}\{A_1 \cup A_2 \cup \cdots \cup A_n\}$ 可以用类似(7.4)的公式算出;这个計算将在第四章第 1 节进行. 这里我們只注意一点: 不等式(7.3)显然对一般情形都成立. 因此,对任意多个事件 A_1, A_2, $\cdots$ 不等式

$$\mathbf{P}\{A_1 \cup A_2 \cup \cdots\} \leqslant \mathbf{P}\{A_1\} + \mathbf{P}\{A_2\} + \cdots \tag{7.6}$$

成立. 特别,当 A_1, A_2, $\cdots$ 互斥时,我們有

$$\mathbf{P}\{A_1 \cup A_2 \cup \cdots\} = \mathbf{P}\{A_1\} + \mathbf{P}\{A_2\} + \cdots. \tag{7.7}$$

有时称(7.6)为布尔(Boole)不等式.

我們首先考虑一个簡单特殊情形：样本空間只有有限个样本点,比如說有N个，而且每一个样本点的概率都是$1/N$. 在这种情形下,任何一个事件A的概率等于其中的样本点的个数除以N. 在古典文献中，样本空間中的点子叫作“情形”，A中的点子叫作“有利情形”(有利于A). 如果全部的样本点的概率都是一样,则事件A的概率等于有利情况除以一切可能情形. 不幸的是,这种叙述被濫用来作为概率的“定义”. 常常以为每一个有限的样本空間中的全部样本点的概率都是一样的. 其实,并不如此. 例如扔一个不均匀的錢币,样本空間只包含两个点,正面和反面，但是它們可以具有任意的概率p和q,其中 $p+q=1$. 一个新生的嬰儿可能是男孩也可能是女孩,但是实际上它們的概率可以不一样. (6.b)給出了一个进一步的反例. 全部样本点的概率都一样的样本空間的用处几乎全部局限于研究机会游戏和組合分析.

8. 問 題

1.在1，2，3，4，5五个数字中先任意抽取一个,然后在剩下的四个中再抽取一个. 假定这全部20个可能的結果都具有相同的概率,試求有一个奇数在(a)第一次被抽出;(b)在第二次被抽出;(c)二次都被抽到的概率.

2.在例(2.a)的样本空間中,給全部27个点以相同的概率. 利用例(4.d)的符号，对事件$A_1=S_1$, $A_2=S_2$来驗証公式(7.4). S_1S_2包含多少样本点?

3.考虑符号1234的24种可能的排列，并且对每一个排列都賦以概率$\frac{1}{24}$. 令A_i为数字i出現在其自然位置(其中$i=1,2,3,4$)的事件.驗証公式(7.4).

4.扔一个錢币,直到它連續地出現两次相同的結果为止. 設扔n次的每一个可能結果都具有相同的概率$\frac{1}{2^n}$. 試描述这个样本空間,并求出下列事件的概率: (a)实驗在扔第6次之前結束;(b)須要扔偶数次才結束.

5.在例(5.b)的样本空間中，我們对(*)中恰巧包含k个字母的样本点賦以概率$\frac{1}{2^k}$. (换句話說,aa和bb具有概率$\frac{1}{4}$，acb 具有概率 $\frac{1}{8}$，等等.)

(a) 証明(*)中的样本点的概率之和为 1，(**) 中之两个样本点的概率为 0.(b)証明 a 胜的概率为 $\frac{5}{14}$. b 胜的概率也是$\frac{5}{14}$，而 c 胜的概率为$\frac{2}{7}$. (c)在第 k 局或在第 k 局前无法判断誰胜誰負的概率为$\frac{1}{2^{k-1}}$.

6.变更例(5.b)，計算每一局中不分胜負的可能性，給出适当的样本空間. 你将如何定义概率?

7.在問題 3 中，証明 $A_1A_2A_3 \subset A_4$ 和 $A_1A_2A_3' \subset A_4'$.

8.利用例(4.d)的符号証明(a) $S_1S_2D_3 = 0$; (b)$S_1D_2 \subset E_3$; (c)$E_3 - D_2S_1 \supset S_2D_1$.

9.擲两顆骰子. 令 A 为点数的和是奇数的事件，B 为至少出現一个么点的事件. 試描述事件 AB, $A \cup B$, AB'. 如果假定全部 36 个样本点都具有相同的概率，試求 AB, $A \cup B$, AB' 的概率.

10.在例(2.g)中，試叙述下列事件的意义: (a)ABC; (b)$A - AB$; (c)$AB'C$.

11.在例(2.g)中，試驗証 $AC' \subset B$.

12.在桥牌游戏中，令 $N_k(k = 1, 2, 3, 4)$为北家至少有 k 个爱司的事件. 以 S_k, E_k, W_k 分别表示南，东，西各家至少有 k 个爱司. 在事件(a)W_1'; (b)N_2S_2; (c)$N_1'S_1'E_1'$; (d)$W_2 - W_3$; (e)$N_1S_1E_1W_1$; (f)N_3W_1; (g)$(N_2 \cup S_2)E_2$ 中，試問西家有几个爱司?

13.在上題中，試驗証 (a) $S_3 \subset S_2$; (b) $S_3W_2 = 0$; (c) $N_2S_1E_1W_1 = 0$; (d)$N_2S_2 \subset W_1'$; (e)$(N_2 \cup S_2)W_3 = 0$; (f)$W_4 = N_1'S_1'E_1'$.

14.試驗証下列关系式[1]:

(a) $(A \cup B)' = A'B'$;

(b) $(A \cup B) - B = A - AB = AB'$;

(c) $AA = A \cup A = A$;

(d) $(A - AB) \cup B = A \cup B$;

(e) $(A \cup B) - AB = AB' \cup A'B$;

(f) $A' \cup B' = (AB)'$;

(g) $(A \cup B)C = AC \cup BC$.

15.化簡: (a)$(A \cup B)(A \cup B')$; (b)$(A \cup B)(A' \cup B)(A \cup B')$;

---

1) 注意，$(A \cup B)'$ 表示 $A \cup B$ 的补集，与 $A' \cup B'$ 的意义不同. 同样，$(AB)'$ 与 $A'B'$ 的意义亦不相同.

(c) $(A\cup B)(B\cup C)$.

16. 試述下列关系中哪些是正确的，哪些是錯誤的：

(a) $(A\cup B)-C=A\cup(B-C)$;

(b) $ABC=AB(C\cup B)$;

(c) $A\cup B\cup C=A\cup(B-AB)\cup(C-AC)$;

(d) $A\cup B=(A-AB)\cup B$;

(e) $(AB\cup BC\cup CA)\supset ABC$;

(f) $(AB\cup BC\cup CA)\subset(A\cup B\cup C)$;

(g) $(A\cup B)-A=B$;

(h) $AB'C\subset A\cup B$;

(i) $(A\cup B\cup C)'=A'B'C'$;

(j) $(A\cup B)'C=A'C\cup B'C$;

(k) $(A\cup B)'C=A'B'C$;

(l) $(A\cup B)'C=C-C(A\cup B)$.

17. 令 A, B, C 为任意三个事件. 試用 A, B, C 来表达下列事件：

(a) 只有 A 发生；

(b) A 和 B 都发生而 C 不发生；

(c) 所有这三个事件都发生；

(d) A, B, C 中至少有一个事件发生；

(e) 至少有二个事件发生；

(f) 恰有一个事件发生；

(g) 恰有二个事件发生；

(h) 沒有一个事件发生；

(i) 不多于二个事件发生.

18. 任意二个事件之拼 $A\cup B$ 可表示成二个互不相容的事件之和，例 $A\cup B=A\cup(B-AB)$. 試用类似的方式表达三个事件 A, B, C 之拼.

19. 利用問題 18 的結果証明

$$\mathbf{P}\{A\cup B\cup C\}=\mathbf{P}\{A\}+\mathbf{P}\{B\}+\mathbf{P}\{C\}-\mathbf{P}(AB)-$$
$$-\mathbf{P}\{AC\}-\mathbf{P}\{BC\}+\mathbf{P}\{ABC\}.$$

[这是第四章(1.5)的特殊情形.]

第二章　組合分析概要

这一章的目的在于导出一些基本的公式，并发展相应的概率背景．有一定程度的讀者可以不讀这章，而直接閱讀第五章，在那里重新提到第一章里所引出的主要理論綫索．

在簡单的机会游戏、抽样手續、占位及序次等等問題里面，我們通常遇到的情况是：有限样本空間，各个样本点的概率相等．这时，要想計算一个事件 A 的概率，只要把 A 中的点（有利情况）的个数，除以样本点（可能情况）的总数就成了．如果有系統地使用几条法則，就可以使这計算更容易一些，現在我們就来把这些法則給以概括．始終遵循着几个标准的工具，可以导致思維的簡化和节省，我們将要遵循的进程就是这样，而不是在一个一个的特例[1]上考究最簡短的計算方法．

1. 預 备 知 識

对子． 給定 m 个元素 $a_1, a_2, \cdots, a_m$ 及 n 个元素 $b_1, b_2, \cdots, b_n$．从以上两組的每組中取出一个元素配成对子 (a_j, b_k)，一共能配成 mn 个对子．

証． 象乘法表那样，把对子用 m 个横列和 n 个縱行排成一个矩形陣，于是 (a_j, b_k) 恰巧处在第 j 列和第 k 行的交点上．因此，每个对子恰好出現一次，从而需証之結論乃属显然．

例． (a) 桥牌（参閱第一章第1节的脚注）．把4种花色和13个面值当作二組元素来看．每张牌都由它的花色和面值决定．因此，这样的組合共有 $4\cdot 13 = 52$ 种，所以就有52张牌．

(b) “七用灯”．广告上叫作“七用灯”的，是这样一种灯：它

1) 有兴趣的讀者可以在經典的教科书[4]里，找到很多初等組合分析的資料．还可参閱文献[5]，其中包含700个附有完全解答的問題．

有 3 个普通的灯泡，外加一个操纵发光的装置，能使每个灯泡在三个不同程度上发光，但也可以不开动此装置. 这四个可能性，每个都可以和 0 个，1 个，2 个或 3 个灯泡結合起来. 所以一共有 16 个組合，其中之一(即(0, 0))使得灯泡一个也不亮. 剩下还有 15 种方式(不是 7 种)来操纵灯.

多元組. 給定 n_1 个元素 $a_1, a_2, \cdots, a_{n_1}$; n_2 个元素 $b_1, b_2, \cdots, b_{n_2}$, 依此类推到 n_r 个元素 $x_1, x_2, \cdots, x_{n_r}$. 从以上各組的每組中取一个元素来配成多元組 $(a_{j_1}, b_{j_2}, \cdots, x_{j_r})$, 一共可以配成 $n_1\cdots n_r$ 个多元組.

証. 若 $r = 2$, 多元組的問題归为对子的問題，因而我們的命題成立. 若 $r = 3$，可取对子(a_i, b_j)作为一类新的元素，則有 n_1n_2 个对子和 n_3 个元素 c_k，每个 3 元組 (a_i, b_j, c_k) 本身可以看作由 (a_i, b_j) 和元素 c_k 所組成的一个对子，因此 3 元組的个数共有 $n_1n_2n_3$ 个. 用归納法可以証明，这个結論对每个 r 都成立.

也許如下所述的是最簡单而且最有用的来描述上面的定理的方法. 为了构成一个多元組$(a_{j_1}, b_{j_2}, \cdots, x_{j_r})$，我們曾經选取一个 a，一个 b，等等. 我們总共选了 r 次，而且依次地有 $n_1, n_2, \cdots, n_r$ 种可能的选法. 由此断言：这种手法会得到 $n_1 \cdot n_2\cdots n_r$ 种不同的結果.

例. (c) 多元分組. 假定把人按其姓别，婚否和职业进行分組，各种状况就相当于元素. 如果社会上只有 17 种职业，則总共可以分成 $2 \cdot 2 \cdot 17 = 68$ 組.

(d) 在一个农业实驗中，要試驗三种不同的处理(如施肥，噴雾剂，温度). 如果这些处理分别有 r_1, r_2 和 r_3 种实施的水平或浓度，則实施的方案共有 $r_1r_2r_3$ 种.

(e) "把球置入盒中"归結为对每一个球选取一个盒. 对于 r 个球，我們就有 r 次独立的选取，因此 r 个球置入 n 个盒中有n^r 种不同的方法. 回忆第一章例(2.b)使我們得知：有很多实驗抽象地看来和"球置入盒"是等价的. 例如，把骰子的各个面都看作"盒子"，則前面的命題断言：把一个骰子擲 r 次共有 6^r 个可能的結

果，其中有 5^r 个結果满足"么点从未出現"的条件．假定全部可能的結果都是等同可能的，"因而 r 次抛擲中不出現"么点的概率为 $\left(\frac{5}{6}\right)^r$．我們可能会这样天眞的期望：擲六次骰子，么点一定出現，其实不然，这一事件的概率仅仅是 $1-\left(\frac{5}{6}\right)^6$，或者小于 2/3 [参看例(3.b)]．

2. 有序样本

考虑 n 个元素 $a_1, a_2, \cdots, a_n$ 所組成的集合或总体．任何 r 个元素 $a_{j_1}, a_{j_2}, \cdots, a_{j_r}$ 的有序排列，就称为从总体中取出的，大小为 r 的一个样本．为了形象化起見，我們可以設想，元素是逐个逐个地抽取的．于是就有两种可能的手續．第 1 种是有放回的抽样；在这里，每个元素的抽取都是在整个总体中进行的，所以同样的元素可以被抽到一次以上．因此在样本的排列中容許有重复的元素．第 2 种是无放回的抽样，在此，某个元素一旦被抽取，即被从总体中除去，所以样本的排列中，不会有重复．显然，在这种情况，样本的大小，不能超过总体的大小 n．

在有放回的抽样里，样本的 r 个元素的每一个都可以有 n 种抽取的方式，所以可能的样本数为 n^r．这一点从刚才的定理可知，只要让定理中的 $n_1 = n_2 = \cdots = n$ 即可．在无放回的抽样里，样本的第 1 个元素可以有 n 种抽取方式，而第 2 个则只有 $(n-1)$ 种，第 3 个只有 $(n-2)$ 种，依此类推．根据上述定理可知，无放回抽样的样本总数是 $n(n-1)\cdots(n-r+1)$．这个乘积經常出現，为了方便起見，引进記号[1)]

$$(n)_r = n(n-1)\cdots(n-r+1). \tag{2.1}$$

显然，当 $r > n$ 时，$(n)_r = 0$．于是我們有下面的

定理． 从 n 个元素所成的总体中，取大小为 r 的抽样．如果抽样是有放回的，則有 n^r 个不同的样本；如果抽样是无放回的，則

1) 記号 $(n)_r$ 并不是标准的，甚至 n 不是整数时，本书也一貫地用它．

有$(n)_r$个不同的样本.

我們注意 $r=n$ 的特别情形，在无放回的抽样中，一个大小为 n 的样本包含了其总体的全部元素，它构成了其全部元素的一个排列. 因此，n 个元素 $a_1, \cdots, a_n$ 有 $(n)_n = n\cdot(n-1)\cdots 2\cdot 1$ 种不同的方式排成次序. 我們用符号 $n!$ 来代替$(n)_n$，$n!$是一个常用的符号. 因此，我們有

系. n 个元素能排成的不同次序的数目为

$$n! = n\cdot(n-1)\cdot\cdots 2\cdot 1.$$

史密斯夫妇是从人类的总体里所抽出的大小为 2 的一个样本，他們也是一切配偶所成的总体里所抽出的大小为 1 的一个样本. 这个例子說明，样本的大小，只有相对于所給定的总体来說才是确定的. 又如把一个錢币扔 r 次，就是从 H 和 T 两个字所成的总体中抽取大小为 r 的一个样本. 这个由 H 和 T（一共 r 个）組成的排列，也就是样本空間（即是对应于扔錢币 r 次的实驗的那个空間）中的一个样本点.

从大小为 n 的总体中抽取 r 个元素是一个实驗，实驗的可能結果就是这些大小为 r 的样本. 它們的总数是 n^r 或 $(n)_r$，随着抽取的方式是否有放回而定. 无論哪一种情况，我們的理想实驗是用样本空間来描述的，其中每一个个别的点都代表大小为 r 的一个样本.

到此为止，我們还沒有談到过联系着这些样本的概率. 我們通常賦予各个样本以相等的概率，这样的样本就叫作随机样本. “随机”这个詞在其一般的意义上还沒有定义，但是用在抽样或选择的場合时，它倒是有着特定的意义的. 无論何时，只要我們談到固定大小 r 的随机样本时，形容詞“随机的”意味着所有可能的样本都具有相同的概率. 在有放回抽样中，这个相同的概率就是 n^{-r}，而在无放回抽样中，则为$\frac{1}{(n)_r}$，而 n 表示取样本的那个总体的大小. 如果 n 大而 r 相对地小，则比值$\frac{(n)_r}{n^r}$接近于 1. 这就說明

了这样的一个事实：如果总体大而样本相对地小，有无放回的二种抽样方式的区别是可以忽略的[参看 11.1，11.2 和第六章 35]。

上面已經引进了应用的术語“随机的”，但是尚未說明我們的随机抽样的模型对于現实世界的适用性。扔錢币、擲骰子和类似的一些游戏，都可以解释为有放回的随机抽样的实驗，并且我們所賦給的概率很接近于在长期持續实驗下所观察到的頻率，即使完全均匀的錢币和骰子并不存在。从一付洗得很好(比平常洗得还要好)的紙牌中，不断地进行抽取，就是无放回抽样的一个典型例子。在人口的抽样中，統計学家往往遇到巨大的和无法預料的困难；經驗表明，在这种場合，甚至連随机性的粗略的形象都是不容易获得的。

練习. 在无放回抽样中，总体中任何一个固定的元素包含在大小为 r 的随机样本中的概率为

$$1-(n-1)_r\div(n)_r=1-\frac{n-r}{n}=\frac{r}{n}.$$

在有放回的抽样中，其对应的概率为

$$1-\{(n-1)/n\}^r.$$

3. 例

我們考虑从一个包含 n 个元素 $a_1,\cdots,a_n$ 的总体中抽出一个大小为 r 的有放回的随机样本。我們感兴趣的是“样本 $(a_{j_1},\cdots,a_{j_r})$ 中沒有重复元素”的事件 A，也就是，这种样本可以从无放回的抽样所得到。前面的定理証明了：不同的样本的总数是 n^r 个，其中有 $(n)_r$ 个满足上述条件。假定各种排列都是等可能的，則我們断言：在我們的样本中沒有重复的概率为

$$p=\frac{(n)_r}{n^r}=\frac{n(n-1)\cdots(n-r+1)}{n^r}. \tag{3.1}$$

下面关于这个公式的具体解释将显示出这个公式的广泛应用。

(a) 随机抽样数。令总体由数字 0，1，$\cdots$，9 所构成。每次抽 5 个数字构成的排列是一个大小为 $r=5$ 的样本，我們假定每一种排列的概率都是 10^{-5}。由 (3.1) 5 个連續随机数字都不相

同的概率为 $p = (10)_5 10^{-5} = 0.3024$.

根据我們直观的想法，在具有小数位很多的大多数数学表格里，最末 5 位数字具有很大的随机性（在普通的对数表和其它一些小表里，表差近似地为常数，因此最末一位小数是有規則地变动的）．作为一个实驗，选取有 16 个位数的表[6]，并把表中所列的那些数——其最末的 5 位数字都不相同的——枚举出来．在每組有 100 个数碼的最初 12 个組中，具有 5 位不同的数字的数碼的个数变动如下：30，27，30，34，26，32，37，36，26，31，36，32．小子样理論証明，这样起伏的量会很好地落在期望限度以內． 这里的平均頻率为 0.3142，与理論上的概率 0.3024 是十分近似的［参閱第七章例(3.f)］．

下面，我們来考虑数 $e = 2.71828\cdots$．把最初的 800 位小数[7]分成 160 段，每段有 5 个数字；并以每組 10 段排成 16 組．在这16組中，所有 5 个数字都不相同的列举如下：

3，1，3，4，4，1，4，4，4，2，3，1，5，4，6，3．

頻率的波动应該在 0.3024 周围． 据小子样拟合理論，上下波动的振幅并不比所期望的大． 在这 160 段里，我們所考虑的事件的整个頻率为 $\frac{52}{160} = 0.325$． 它很合理地近似于 $p = 0.3024$．

(b) 如果 n 个球随机地置入 n 个盒中，每一个盒都放球的概率为 $n!/n^n$． 这个概率非常小：对 $n=7$ 来說，它为 $0.00612\cdots$．这就意味着如果在一个城市中每一周发生 7 次意外的事件，則（假定各种分布都是等可能的）实际上每一周都包含有这样一些天，它发生两件或两件以上的意外事件． 平均来說，大約 165 周才有一周每一天都发生一件意外的事件．这个例子說明了随机性的一个意料不到的性質．（7 个球置入 7 个盒中的全部可能情形在第 5 节表 1 中都列出了，两个或两个以上的盒空着的概率大約为 0.87.）当 $n = 6$ 时，概率 $n!\, n^{-n}$ 等于 $0.01543\cdots$．这說明把一个均匀骰子擲六次，要得到每一次出現的点数都不相同是如何的不容易．［一个給定的点不出現的概率大約为 $\frac{1}{3}$；参看例(1.e).］

(c) 电梯. 在开始时載有 7 位乘客的一架电梯，在 $n = 10$ 层的楼房的每一层上都停留，試問不发生 2 位乘客在同一层楼上离开电梯的概率 p 为何? 为了使問題确定起見，假定乘客离开电梯的各种安排有同等的概率(这只是大概的,粗略的近似). 于是

$$p = 10^{-7}(10)_7 = (10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4)10^{-7} = 0.06048.$$

假使这事件一旦发生,我們就認为发生这件事是奇特的,而对重复来說发生的可能性在 1000 次中只一次(参閱問題 10.43 的答案).

(d) 生日. 任意 r 个人的生日可以看成为从一年的全部日子所构成的总体中抽取的大小为 r 的一个样本. 虽然每年的时間不是等长的,而在整个一年中,人的出生率也不是固定不变的，但是,作为第一步近似,我們还認为,随机地抽取人,和随机地抽取生日是一样的. 其次,为了簡单起見,每年就以 365 日来考虑.

有了这种約定利用(3.1)，我們算出所有 r 个生日都不相同的概率 p 为[8]

$$p = \frac{(365)_r}{365^r} = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right)\left(1 - \frac{3}{365}\right)\cdots\left(1 - \frac{r-1}{365}\right). \tag{3.2}$$

其数值結果是令人吃惊的. 对 $r = 23$ 个人来說，$p < \frac{1}{2}$，也就是說,对 23 个人来說,至少有两人同一天生日的概率超过 $\frac{1}{2}$.

公式(3.2)看来十分繁笨，但是可以容易地推出近似表达式. 如果 r 小,各因子的第二項之間的乘积都可以忽略,于是有一个粗略的近似公式:

$$p \approx 1 - \frac{1 + 2 + \cdots + (r-1)}{365} = 1 - \frac{r(r-1)}{730}. \tag{3.3}$$

例如当 $r = 10$ 时,正确的数值为 $p = 0.883\cdots$，而(3.3)給出的近似值为 0.877.

对于較大的 r 可以用对数法求得很好的近似值. 只要 x 为正且很小,有 $\log(1-x) \approx -x$. 于是由(3.2)可得

$$\log p \approx -\frac{1+2+\cdots+(r-1)}{365} = -\frac{r(r-1)}{730}. \quad (3.4)$$

对于 $r = 30$，这个公式給出的近似值为 0.3037，而正确值为 $p=0.294$. 对于 $r \leqslant 40$，(3.4)的誤差小于 0.08.（在第 7 节中継續討論此問題，也可参看問題 10.44 的答案.）

4. 子体和分划

和以前一样，我們用大小为 n 的总体来表示一个由 n 个元素构成的不分次序的集合. 所謂两个总体不一样，意思就是說有一个总体有一个元素不属于另一个总体. 从一个大小为 n 的总体里选取 r 个元素就构成了一个大小为 r 的子体. 这种选法有多少种呢？每一个大小为 r 的子体可以排列成 $r!$ 种不同的次序，因此用这种办法可以产生 $r!$ 个不同的沒有重复的样本. 反之，每一个这样的大小为 r 的样本包含 r 个不同的元素，因此它确定了一个大小为 r 的子体. 我們知道上述的样本一共有 $(n)_r$ 个. 如果 x 是大小为 r 的子体的个数，则有序样本的个数为 $x \cdot r!$，因此我們断言 $x = (n)_r/r!$. 这种形式的数目称为二項系数，它們的标准符号是

$$\binom{n}{r} = \frac{(n)_r}{r!} = \frac{n(n-1)\cdots(n-r+1)}{1\cdot 2\cdots(r-1)\cdot r}. \quad (4.1)$$

因此，我們証明了

定理 1. 一个包含 n 个元素的总体可以产生 $\binom{n}{r}$ 个不同的子体($r \leqslant n$).

換句話說，从 n 个元素中可以用 $\binom{n}{r}$ 种不同的方法选出包含 r 个元素的子羣. 因为从 n 个元素的总体中选取 r 个元素和留下 $n-r$ 个是等价的，因此，对 $r \leqslant n$ 显然有

$$\binom{n}{r} = \binom{n}{n-r}. \quad (4.2)$$

为了直接証明方程(4.2)，我們只須注意：把二項系数(4.1)換成下面的写法

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}. \tag{4.3}$$

[只需把(4.1)的分子分母同时乘以$(n-r)!$即得(4.3).]注意：方程(4.2)的左边当 $r=0$ 时是无定义的,但右边却是有定义的. 为了使(4.2)对一切满足 $0 \leqslant r \leqslant n$ 的整数 r 都成立,我們定义

$$\binom{n}{0} = 1, \quad 0! = 1, \tag{4.4}$$

和$(n)_0 = 1$.

例. (a) 桥牌和扑克（参看第一章脚注(1)）. 因为在一个人的手中的牌的次序是互不相干的，前面的定理証明了一付紙牌有$\binom{52}{13} = 635,013,559,600$ 种不同的桥牌和$\binom{52}{5} = 2,598,963$种扑克. 試我們来計算一个人的手里的扑克包含 5 张不同面值的概率 x. 这些面值可以用$\binom{13}{5}$种方法选取,而且对应于每一张牌我們可以自由地由四种花色中的一种选取. 由此推出，$x=4^5 \cdot \binom{13}{5} \div \binom{52}{5}$,它近似地为 0.5071. 在一个人的手里的桥牌中，13 张面值都不同的概率为 $4^{13} \div \binom{52}{13}$,或者近似地为 0.0001057.

(b) 48 个州中每州有 2 个参議員. 考虑由其中任意选取 48 个参議員組成委員会这样的事件. 假設 (1) 指定某州要有代表；(2)所有的州都要有代表.

在第一种情形,其逆事件的概率 q 要好算些. 所謂逆事件,就是指定的那个州沒有一个代表作为委員. 总共 96 个代表中,有94 个不是那个州的代表. 因此

$$q = \binom{94}{48} \div \binom{96}{48} = \frac{48 \cdot 47}{96 \cdot 95} = 0.24737\cdots.$$

其次，第 2 节的定理証明：委員会要使每一州都有一个代表的选取方式共有 2^{48}种.因此,所有的州都有委員的概率为 $p=2^{48} \div \binom{96}{48}$. 利用史特令公式（参看第 9 节），可証 $p \approx (3\pi)^{\frac{1}{2}} 2^{-46} \approx 4 \cdot 10^{-14}$.

(c) 占位問題. 我們再一次考慮 r 个球随机地分布于 n 个盒中(即是 n^r 种可能的排列都具有概率 n^{-r}). 为了找出一个給定的盒中恰巧包含 k 个球($k=0, 1, 2, \cdots$)的概率 p_k, 我們首先注意: k 个球可以有 $\binom{r}{k}$ 种方式选取, 而剩下的 $r-k$ 个球放入剩下的 $n-1$ 个盒的方式有 $(n-1)^{r-k}$ 种. 由此推出

$$p_k = \binom{r}{k} \cdot \frac{1}{n^r} \cdot (n-1)^{r-k} = \binom{r}{k} \cdot \frac{1}{n^k} \cdot \left(1 - \frac{1}{n}\right)^{r-k}. \quad (4.5)$$

这就是所謂二項分布的一种特殊情形, 二項分布在第六章中將要討論. 数值可以从第六章表 3 中查出.

(d) 包含两类元素的次序. 考虑一个由 $n = a + b$ 个元素构成的总体, 其中 a 个是属于同一类的, 而其余 b 个是属于另一类的. 为了方便起見, 我們用 $\alpha_1, \cdots, \alpha_a$ 和 $\beta_1, \cdots, \beta_b$ 代表其元素. 这些元素可以排成 $n!$ 种不同的次序. 然而, 如果我們約定所有的 α 和所有的 β 內部是不可分辨的(即是不考虑其指标), 那末, 某些次序就变为不可分辨的了. 事实上, 每一个次序由其 a 个 α 的位置所唯一确定, 而这些 α 可以由 $\binom{a+b}{a} = \binom{a+b}{b}$ 种不同的方式选取. 因此, 一个由 a 个不可辨别的 α 和 b 个不可辨别的 β 所构成的总体可以排列成 $\binom{a+b}{a} = \binom{a+b}{b}$ 种可以辨别的次序. (例如, 序 $\alpha\alpha\alpha\beta\beta$ 可以排列成 10 个可以辨别的次序.) α 中的任何一种排列或者 β 中的任何一种排列都不改变其外貌, 所以有 $a!b!$ 种排列, 其外貌是一样的. 由此推出, 如果我們賦予 $(a+b)!$ 种排列中的每一种以概率 $1/(a+b)!$, 則全部可辨别的排列都是等可能的, 而且每一个都有概率 $a!b! \div (a+b)!$. 因此, 如果我們說到关于等可能的排列, 它既代表可辨别的排列又代表这些元素的全部排列的总体. (可把它和球随机放入盒中的情形作比較——参看第 5 节.)

定理 2. 令 $r_1, \cdots, r_k$ 为滿足

$$r_1 + r_2 + \cdots + r_k = n, \quad r_i \geq 0 \quad (4.6)$$

的整数. 把一个包含 n 个元素的总体分成 k 个有序的部分(分为 k 个子体), 其中第一部分包含 r_1 个元素, 第二部分包含 r_2 个元素,等等,这种分法共有

$$\frac{n!}{r_1!\,r_2!\cdots r_k!}. \tag{4.7}$$

(数(4.7)叫作多項系数.)

[注意: 子体之間的次序是必要的, 譬如 $(r_1=2, r_2=3)$ 和 $(r_1=3, r_2=2)$ 代表不同的分法; 然而,組內之間的次序是无关紧要的. 同时还要注意: $0!=1$, 所以 r_i 等于 0 时并不影响公式(4.7).]

証. 反复利用(4.3)可以証明数(4.7)能够写成下述形式:

$$\binom{n}{r_1}\binom{n-r_1}{r_2}\binom{n-r_1-r_2}{r_3}\cdots\binom{n-r_1-\cdots-r_{k-2}}{r_{k-1}}. \tag{4.8}$$

另一方面,为了作成需要的分划,我們首先从給定的 n 个元素中选取 r_1 个; 再从剩下的 $n-r_1$ 中选取大小为 r_2 的第二組,等等. 当第 $k-1$ 組选出以后,还剩下 $n-r_1-r_2-\cdots-r_{k-1}=r_k$ 个元素,而这些元素就构成最后一組. 我們断言,构成上述分划的办法共有(4.8)所确定的数目那样多.

例. (e) 桥牌. 在一場桥牌中, 52 张牌分成四組, 每組 13 张, 因此, 不同的情形共有 $52!(13!)^{-4}=(5\cdot 36\cdots)\cdot 10^{28}$. 让我們計算每一个参与者都有一张爱司的概率. 4 张爱司可以排成 $4!=24$种不同的排列次序,而每一种次序代表給每个参与者一张爱司. 剩下的 48 张牌有$(48!)(12!)^{-4}$ 种分配方式. 因此所要求的概率为

$$24\cdot 48!\cdot(13)^4\div 52!=0.105\cdots.$$

(f) 骰子. 把 12 顆骰子擲一次, 可以产生 6^{12} 个不同的結果,它們都被賦以相同的概率. 每点出現二次的事件,即把 12 个骰子按排成 6 組,每組 2 个. 因此,这事件的概率为 $12!/(2^6\cdot 6^{12})=0.003438\cdots$.

(在定理 2 中,可以允許 $r_i=0$, 因此实际上 n 个元素常分为

k 个或更少的子体. $r_i > 0$ 即恰巧分化为 k 类的情形将在問題 11.7 处理.)

*5. 在占位問題中的应用

第一章第 2 节的例子曾指出把 r 个球随机地放入 n 个盒中的模型具有广泛的用处. 我們再回过来討論这个模型,当然,我們假定这 n^r 种可能的分布中的每一种分布的概率都是 n^{-r}. 一个特定的分布的最重要的性質可以由它的占位数目 $r_1, \cdots, r_n$ 来表示,其中 r_i 是放入第 i 个盒中的球的个数. 其中

$$r_1 + r_2 + \cdots + r_n = r, \qquad r_i \geqslant 0. \tag{5.1}$$

我們只研究球是不可辨的情形. 因此,球的分布完全由它的占位数来描述,两个分布只有当对应的有序的 n 維向量$(r_1, \cdots, r_n)$不相等时才可以区别. 我們第一个目的是証明

引理. 可区别的分布[即是方程(5.1)的不同的解]的个数为[1]

$$A_{r,n} = \binom{n+r-1}{r} = \binom{n+r-1}{n-1}. \tag{5.2}$$

沒有一个盒是空的可区别的分布的个数共有

$$\binom{r-1}{n-1}.$$

証. 我們用 $n+1$ 条 | 把空間分成 n 个盒,而用 * 代表球. 因此符号 | * * * | * | | | | * * * * | 代表这样一个分布: $r = 8$ 个球放于 $n = 6$ 个盒中,其占位数为 3, 1, 0, 0, 0, 4. 这样的符号的开始和終結必需有一条 |,但是其余的 $n-1$ 条 | 和 r 个 * 可以按任意次序出現. 用这种方法可以看出,可区别的分布数等于从 $n+r-1$ 个位置中任取 r 个位置的选法. 沒有一个盒是空的相当于沒有两条 | 靠在一起. 但 r 个星之間空出 $r-1$ 个位置,而这 $r-1$ 个位置中有 $n-1$ 个被 | 所占位, 因此,我們有$\binom{r-1}{n-1}$种可能的选择,这就証明了我們的引理.

---

* 这一节的內容是很有用的,而且是富有启发性的,不过以后不是明显地用到.

1) $r = 100$, $n = 4$ 的特例在第一章例(2.e)中应用过.

例. (a) 擲 r 个不能辨别的骰子，可以分辨的結果共有 $\binom{r+5}{5}$ 个.

(b) 偏导数. 一个 n 个变元的解析函数 $f(x_1, \cdots, x_n)$ 的 r 阶偏导数不依賴其求导数的次序而只依賴于对每一个变元求导数的次数. 因此，每一个变元对应于一个盒，所以不同的 r 阶偏导数共有 $\binom{n+r-1}{r}$ 个. 一个三元函数的 4 阶偏导数共有 15 个，5 阶偏导数共有 21 个.

把 r 个球放入 n 个盒中是对 r 个球的总体的一种分法. 由第 4 节定理 2 可得知：具有占位数 $r_1, \cdots, r_n$ 的分布共有 $r! \div (r_1! \cdot r_2! \cdots r_n!)$. 这个公式仍然包含了占位数的次序关系，但是，这种次序常常是非本質的. 下面的例子就是想用来說明解决許多初等的組合問題的簡单而常用的方法.

例. (c) 把 $r=7$ 个球放入 $n=7$ 个盒中的形式. (这些盒可以理解为一周的 7 天，这些球可以理解为呼喚，信件，意外的事件，等等.) 为了簡洁起見，讓我們考慮一个以任意次序出現的占位数为 2, 2, 1, 1, 1, 0, 0 的分布. 这 7 个占位数把 7 个盒分为三个子类，它們分别为有两个盒两次地被占位；三个盒被占位一次；两个空着的盒. 这样把它們分为三羣其大小各为 2, 3, 2 的分法共有 $7! \div (2! \cdot 3! \cdot 2!)$ 种. 而对每一种給定的分法，把 7 个球放入这 7 个盒中的方法共有 $7! \div (2! \cdot 2! \cdot 1! \cdot 1! \cdot 1! \cdot 0! \cdot 0!) = 7! \div (2! \cdot 2!)$ 个不同的分布. 所以用某种次序出現的占位数为 2, 2, 1, 1, 1, 0, 0 的不同的分布共有

$$\frac{7!}{2! \cdot 3! \cdot 2!} \times \frac{7!}{2! \cdot 2!}. \qquad (5.3)$$

注意：这个結果曾經由(4.7)的双重应用(既对球，又对盒)而推出过. 同样的結果可以用很多种方法推出和表述，不过現在这种方法对很多种类型的問題提供了一种最簡单最常用的技巧（参看第 10 节問題 43—45). 表 1 包含类似 (5.3) 的內容，而且还包含在 $r=n=7$ 的情形下各种占位数的一切可能形式的概率.

表 1. 7 个球放入 7 个盒中的随机分布

占位数	排列数等于 7!×7!除以	概率(排列数除以7^7)
1,1,1,1,1,1,1	7!×1!	0.006 120
2,1,1,1,1,1,0	5!×2!	.128 518
2,2,1,1,1,0,0	2!3!2!×2!2!	.321 295
2,2,2,1,0,0,0	3!3!×2!2!2!	.107 098
3,1,1,1,1,0,0	4!2!×3!	.107 098
3,2,1,1,0,0,0	2!3!×3!2!	.214 197
3,2,2,0,0,0,0	2!4!×3!2!2!	.026 775
3,3,1,0,0,0,0	2!4!×3!3!	.017 850
4,1,1,1,0,0,0	3!3!×4!	.035 699
4,2,1,0,0,0,0	4!×4!2!	.026 775
4,3,0,0,0,0,0	5!×4!3!	.001 785
5,1,1,0,0,0,0	2!4!×5!	.005 355
5,2,0,0,0,0,0	5!×5!2!	.001 071
6,1,0,0,0,0,0	5!×6!	.000 357
7,0,0,0,0,0,0	6!×7!	.000 008

关于波司-爱因斯坦和弗米-迪拉克統計的注. 到現在为止，我們一直假定 n^r 种可能分布中每一个可能分布的概率都是 n^{-r}. 事实和經驗都迫使物理学家抛弃这种假定,而以各种不同的方法賦以概率.

考虑一个力学系統，它具有 r 个不能分辨的質点. 在統計力学中，常常把相空間分成 n 个小区域（或者盒），n 是很大的，因此，每一个分子总处于某一个盒中. 用这种方法,整个系統的情况以 r 个分子放入 n 个盒的随机分布来描述. 乍一看来,似乎(至少对 n 个盒适当的定义)全部 n^r 个排列都具有相同的概率. 如果这是对的,那末,物理学家把它叫作馬克司威尔-波茨曼統計(这里所用的“統計”一詞是物理中的术語). 人們曾經花費过很多力量要想証明物理質点是按照馬克司威尔-波茨曼的統計法則行动的，但是現代理論无可置疑地表明,这个統計法則对于知識范围內所知的任何質点都是不适用的;在何任情形下,那 n^r 种排列的概率都不是近乎相等的. 于是,人們另外引进两个概率模型. 这两个概率模型各自地能够滿意地描述某一类型的質点的动态. 应用的时候,哪个成功就用哪个. 两个模型都不是放之四海而皆准的. 說不定有一天,对于某些种类的質点人們还会引进新的模型.

必須記住,現在我們只討論那些不可分辨的質点. 設有 r 个質点和 n 个

盒. 所謂波司-爱因斯坦統計法則,就是只考慮那些可区分的排列,而每个排列都賦以相等的概率

$$\binom{n+r-1}{r}^{-1}. \tag{5.4}$$

在統計力学中,已經証明这个假設对于含有偶数个基本質点的光子、核子和原子都是正确的(参看[9]). 为了描述另外一些質点的情况,我們要引进第三种賦概的方法. 弗米-迪拉克統計是基于这些假設: (1)2个或多个質点在同一个盒中是不可能的; (2)满足上一个条件的所有的可区分的排列具有相同的概率. 这里第一个假定要求 $r \leq n$. 每一种排列都相当于 n 个盒中某 r 个各含一个質点,也就是說,每一种安排相当于从 n 个盒中选取 r 个. 因此,共有 $\binom{n}{r}$ 种不同的安排. 根据第二个假設,每一种排列都具有概率 $\binom{n}{r}^{-1}$. 这个模型可以应用在电子、中子和質子上. 这个富有教育性的实例告訴我們,光凭先驗的論証来选择或判明概率模型是行不通的. 事实上,沒有任何的抽象推理可以說明为什么光子和質子并不遵从同一个概率規律. (馬克司威尔-波茨曼統計法則和弗米-迪拉克統計法則之間的基本区别将在第11节問題14—19中闡明.)

在馬克司威尔-波茨曼統計中,把盒 $1, 2, \cdots, n$ 各包含 $r_1, r_2, \cdots, r_n$ 个球的概率加起来(其中 $r_1 + \cdots + r_n = r$)就等于

$$\frac{r!}{r_1! \cdot r_2! \cdots r_n!} n^{-r}, \tag{5.5}$$

在波司-爱因斯坦統計中, 这个概率由(5.4)所給出; 而在弗米-迪拉克統計中,这个概率等于 $\binom{n}{r}^{-1}$,其中每一个 r_i 都为0或为1.注意,"馬克司威尔-波茨曼統計"是物理学家的語言,我們把它叫作球放入盒的随机排列.

例. (a) 令 $n = 5$, $r = 3$. 按照馬克司威尔-波茨曼、波司-爱因斯坦、弗米-迪拉克三种不同的統計法則, 排列(*|—|*|*|—)分别具有概率 $\frac{6}{125}$, $\frac{1}{35}$ 或 $\frac{1}{10}$. 也可以参看第一章例(6.b).

(b) 印错. 一本书中包含了 n 个符号(字母),其中 r 个是印错的. 错字的分布对应于 r 个球放入 n 个盒的分布(每一个盒不可能有多于一个球). 因此,有理由假定: 错字遵从弗米-迪拉克統計法則(参看問題 10.38).

***5a. 应用到連貫中去.** 二类元素按任意次序混合排成一列, 其中每个由同类元素所組成的不能再长的小段称为一个連貫. 例如,在序列 $\alpha\alpha\alpha\beta\alpha x$

$\beta\beta\beta\alpha$ 中，开始时，有长度为 3 的一个 α-連貫，其后有四个連貫，长度依次为 1, 2, 3, 1. α-連貫和 β-連貫是相間的，因此，連貫的个数总是等于异类相邻的总次数加 1.

应用的例子. 連貫理論可以有很多方式应用到統計中去，不过，其主要用途还是在于与随机性检驗和齐次性检驗之間的联系.

(a) 在随机性检驗中，問題是要判断一次給定的观察是不是随机的，或者是不是有系統的因素在起作用. 作为一个簡单的例子[10]，我們假定一个人观察餐桌用餐情况. 他用 E 表示空座，O 表示被占座，給出了如下的一个記录: $EOEEOEEEOEEEOEOE$. 值得注意的是，沒有 2 个被占座位是相邻的. 这可能是偶然的嗎? 因为对于 5 个被占座位和 11 个空座而言，連貫的个数不可能多于 11 个，而这个記录恰巧就是有 11 个連貫的极端情形. 以后，我們会指出，如果所有的排列都是等概的話，出現11个連貫的概率只有0.0578…. 这个小概率在某种程度上符合了这种預感，那就是: 所观察到的那种分离現象是有意的. 这样的怀疑虽然不能用統計的法則加以証明，但是如果継續观察下去，就可以收集到更多的証据. 如果餐厅常常有家庭顾客用餐，那末将会有成羣地集中占領座位的傾向，因而会出現个数(相对地)較少的連貫. 同样，如果把教室里男孩和女孩的連貫計算一下，就也許可以发覚他們混合的情况不是随机的. 难得出現的排列竟然出現，可以給出一些綫索去寻求系統原因: 連貫的数目过多就表示人为的故意混合; 連貫个数太少表示人为的故意集結. 当然，这些結論并不是很可靠地成立的. 但是，有效的統計技巧已經大大地发展了，它在实践中大大地縮小了作出錯誤結論的危险性.

休哈特把連貫理論应用于工业上的质量控制中. 当橡皮垫圈制成之后，其厚度各不相同. 厚垫圈的长串(长連貫)的出現，可能反映生产过程不够完善，因此，引导我們去消除不完善的原因. 这样就能防止即将到来的毛病，而使得产品达到高度的均勻性.

在生物的現場实驗中，我們計算健全作物和有病作物的連串，有病作物的长連貫提示疾病的蔓延. 气象学家留心干燥和潮湿月份的連續現象[11]，借以发現气候的持續現象的綫索.

(b) 为了了解典型的齐次性問題，我們假定两种葯給两組病人服用，或者說我們想比較两种处理(医葯上的，农业上的或工业上的)的效果. 在实践中，我們有两組观察結果，譬如說 $\alpha_1, \alpha_2, \cdots, \alpha_a$ 和 $\beta_1, \beta_2, \cdots, \beta_b$ 分別对应于两种处理或者說分別对应于两个总体中的元素的某一特性(例如重量). α 和 β 都是数，我們假定它按大小次序排列: $\alpha_1 \leqslant \alpha_2 \leqslant \cdots \leqslant \alpha_a$; $\beta_1 \leqslant$

$\beta_2 \leqslant \cdots \leqslant \beta_b$. 我們把这两組数合在一起，并按大小次序来排列. 一个极端的情况是全部 α 都在全部 β 前面，这种現象可以用来指出这两种处理或这两个总体之間有差异. 另一方面，如果两种处理都是一样的，則 α 和 β 或多或少以随机次序出現. 瓦尔德(A. Wald)和华尔夫維茨[12]曾經証明了：連貫理論常常能很方便的利用来发現微小的系統誤差.（一个用不同的方法处理的例子見于第三章第一节.）

許多关于連貫的問題都可以用一种极簡单的办法解决. 給定 a 个不可区分的 α 和 b 个不可区分的 β，則有 $\binom{a+b}{a}$ 个可以区分的次序. 如果有 n_1 个 α-連貫，則 β-連貫的个数或則为 n_1+1，或則为 n_1-1，或則为 n_1. 把 a 个 α 安排在 n_1 个連貫中等价于把它們安排到 n_1 个盒中去，它們中間沒有一个是空的. 由最后一个引理得知：这种安排法共有 $\binom{a-1}{n_1-1}$ 种. 由此推出，具有 n_1 个 α-連貫，n_1+1 个 β-連貫的排列共有 $\binom{a-1}{n_1-1}\binom{b-1}{n_1}$（在第 11 节問題 20—25 中还要継續）.

(c) 在物理学中，用連貫理論来研究合作現象，在艾辛(Ising)的一維点陣理論中，能量依賴于“异类相邻陣点”的总数，也就是說依賴于連貫的个数.

6. 超几何分布

很多組合問題可以归結成下面的形式. 在 n 个元素的总体里，n_1 个是紅的，$n_2 = n - n_1$ 个是黑的. 任意抽出一組 r 个元素. 試求所取出的这一組中，恰有 k 个紅元素的概率 q_k. 这里，k 是正整数，它的取值范围是从 0 到 $\min(r, n_1)$.

为了求出 q_k，我們注意到，选取的这組中，包含着 k 个紅元素和 $r-k$ 个黑元素. 紅元素的选取有 $\binom{n_1}{k}$ 个不同方式，而黑元素有 $\binom{n-n_1}{r-k}$ 个方式. 由于 k 个紅元素的任意一种选法都可以和黑元素的任意一种选法配合，所以我們得到

$$q_k = \frac{\binom{n_1}{k}\binom{n-n_1}{r-k}}{\binom{n}{r}}. \tag{6.1}$$

如此定义的这一組概率,就称为超几何分布[1]. 利用(4.3)式，可以把(6.1)写成

$$q_k = \frac{\binom{r}{k}\binom{n-r}{n_1-k}}{\binom{n}{n_1}}. \tag{6.2}$$

注. 概率 q_k 只对 k 的数值不超过 r 或 n_1 而定义的，虽然如此，但从定义(4.1)可以推出： 当 $b > a$ 时 $\binom{a}{b} = 0$. 因此,当 $k > n_1$ 或 $k > r$ 时,(6.1)和(6.2)式都给出 $q_k = 0$. 于是,(6.1)和(6.2)的定义对所有的 $k \geqslant 0$ 都可以应用，只要把关系 $q_k = 0$ 的出現解释为其所对应的事件是不可能事件就成了.

例. (a) 質量检查. 在工业的質量控制中,产品分批进行抽样检查,每批由 n 件产品組成. 在一批中,次品就以"紅"元素代表之. 当然,它們的数目 n_1 是未知的. 在取出的大小为 r 的一个样本中,次品的个数 k 是完全确定的. 于是,(6.1)式可以使我們对 n_1 的值作出概然的估計;这是統計估值的一个典型問題,但它已超出了本书的范围.

(b) 在例(4.b)中，由 $n = 96$ 位参議員組成总体,有 $n_1 = 2$ 位是指定的那个州(即"紅"的)的参議員. 任意选取 $r = 48$ 位为一組,其中包含指定的那个州的代表可以有 $k=0$, 1 或 2 位. 从(6.2)式,我們可以得到$\left(\text{記住由(4.4)式給出的} \binom{n}{0} = 1\right)$

$$q_0 = \frac{48 \cdot 47}{96 \cdot 95} = 0.24737\cdots,$$

$$q_1 = \frac{48}{95} = 0.50527\cdots,$$

$$q_2 = \frac{48 \cdot 47}{96 \cdot 95} = 0.24737\cdots.$$

例(4.b)中也得到过 q_0 的值,但是方法不同.

---

1) 这个名字可由以下事实来作解释: $\{q_k\}$ 的母函数(参閱第十一章)可以用超几何函数来表示.

(c) 用标記后再捕記录来估計一羣动物的总数[1]. 假設从某一个湖里,捉出 1000 尾魚,涂以紅点后放回. 隔了一个时間以后,重新又捉出 1000 尾魚,发現其中有 100 尾涂有紅点的魚. 根据这些情况,对該湖中魚的总数能作出什么結論呢? 这是統計估值的一个典型問題. 如要叙述近代統計中用到的各种方法,未免离题太远. 現在我們只想說明一下,超几何分布是怎样供給我們綫索来解决問題的. 我們自然要假定,这二次捕捉都考虑为从湖中所有的魚組成的总体中所作的随机抽样.(实际上,这个假設排斥了下面的二种情况: 二次捕捉都在某一个局部的区域进行; 二次捕捉之間相隔的时間太短.)我們还假定,在二次捕捉期間,湖中魚的总数是常数.

为了使問題一般化,我們考虑任意大小的样本,令

n = 湖中魚的总数(未知的);

n_1 = 第一次捉出的魚数,它們相应"紅球";

r = 第二次捉出的魚数;

k = 第二次捉出的魚中紅点魚的总数;

$q_k(n)$ = 第二次捉出的魚中恰有 k 尾紅点魚的概率.

这样一来,我們馬上看出,$q_k(n)$ 是由(6.1)給出的. 实际上,n_1, r 和 k 是可以观察到的,而 n 是未知的. 注意,順便再提一下,n 是一个定数,它不是随机遇而改变的. 所以,如果要問 n 大于某数(譬如說 6000)的概率为何,那是毫无意义的. 但是我們知道,曾經有 $n_1 + r - k$ 尾不同的魚被捕到过,因而 $n \geq n_1 + r - k$. 这就是預先可以作出肯定判断的全部. 在我們的例子里,$n_1 = r = 1000$, $k = 100$, 因此湖中只有 1900 尾魚的假定是可以考虑的. 虽然如此,我們假如从这个假定出发,导出来的結論是: 发生的事件的概率小得惊人. 事实上,大小为 1000 的二次样本,取尽了整个总体中的 1900 尾魚,其概率由(6.1)可知为

1) 这个例子在第一版中就采用了,但是当然我不知道这个方法在实际里已經广泛采用. 关于这方面的新的貢献請参看文献 N. T. J. Bailey[13] 和 D. G. Chapman[14]

$$\frac{\binom{1000}{100}\binom{900}{900}}{\binom{1900}{1000}}=\frac{(1000!)^2}{100!\ \ 1900!}.$$

由史特令公式（参閱第 9 节）可以証明，这个概率的数量級将是 10^{-430}. 因此可以认为，我們的假定不合理而加以否定. 同理可知，n 很大（譬如說 1,000,000）的假定也一样会被否定的. 經过这样的考虑，迫使我們去寻求使 $q_k(n)$ 达到最大值的那个 n，由于那个 n 能使我們的观察有着最大的概率. 对任意一組特定的观察值 n_1, r, k，使 $q_k(n)$ 达到最大值的那个 n 就記作 $\hat{n}$，并称为 n 的**最大似然率估值**. 这个概念是由費歇（R. A. Fisher）所引进的. 为了找出 $\hat{n}$，考虑比值

$$\frac{q_k(n)}{q_k(n-1)}=\frac{(n-n_1)(n-r)}{(n-n_1-r+k)n}. \tag{6.3}$$

由簡单的計算可知，这个比值大于或小于 1，分别按照 $nk<n_1r$ 或 $nk>n_1r$ 而定. 这就是說，当 n 增大时，序列 $q_k(n)$ 先是上升而后下降；当 n 为小于 $\frac{n_1r}{k}$ 的最大整数时，它达到最大值. 因此 $\hat{n}$ 大約等于 $\frac{n_1r}{k}$. 在我們的特殊例子里，魚数的最大似然率估值为 $\hat{n}=10,000$.

n 的眞值可能大些或者小些；我們可以找出这样一个范围，使得我們可以合理地期望 n 落在这个范围之內. 为此，让我們来检驗 n 小于 8500 的这个假設. 把 $n=8500$，$n_1=r=1000$ 代入 (6.1) 而进行計算，在第二次样本中含有 100 尾或較少的紅点魚的概率. 这概率为 $x=q_0+q_1+\cdots+q_{100}$. 直接計算这个 x 是很麻煩的，如用第七章的正态近似法，就可以很容易求出 $x\approx0.04$. 同样，設 $n=12000$，則在第二次的样本中含有 100 尾或較多紅点魚的概率約为 0.03. 根据这些数字，可以推想魚的眞正条数 n 大約处在 8,500 与 12,000 之間. 还有其他的方法来推导这些結論和进行估值，但是我們在这里就不細講了.

由概率 q_k 的定义,可得 $q_0+q_1+\cdots=1$. 因此,对任意正整数 n, n_1 和 r, (6.2)蘊含着

$$\binom{r}{0}\binom{n-r}{n_1}+\binom{r}{1}\binom{n-r}{n_1-1}+\cdots+\binom{r}{n_1}\binom{n-r}{0}=\binom{n}{n_1}. \tag{6.4}$$

这恆等式是常常用到的. 对于正整数 n 和 r 来說, 算是把它証明了,但它的成立并不受这个限制. 实际上,当 n 和 r 为任意正、負整数时,(6.4)都成立(n_1 必須是正整数,否则沒有意义). (两个証明的提示在第 12 节問題 8,9 中給出.)

很容易把超几何分布推广到大小为 n 的原来的总体中含有好几类元素的情况. 例如,假設总体中含有三类元素,其个数分别为 n_1, n_2 和 $n-n_1-n_2$. 如果取出一个大小为 r 的样本,那么,这样本含有 k_1 个第一类元素, k_2 个第二类元素和 $r-k_1-k_2$ 个第三类的概率为(与(6.1)相似)

$$\frac{\binom{n_1}{k_1}\binom{n_2}{k_2}\binom{n-n_1-n_2}{r-k_1-k_2}}{\binom{n}{r}}. \tag{6.5}$$

当然,必須要有 $k_1\leqslant n_1$, $k_2\leqslant n_2$ 和 $r-k_1-k_2\leqslant n-n_1-n_2$.

例. (d) 桥牌. 52 张牌的总体包含四个类,每类由十三个元素組成. 任发到一家的 13 张牌含有 5 张黑桃, 4 张紅心, 3 张方块,和 1 张梅花的概率为

$$\frac{\binom{13}{5}\binom{13}{4}\binom{13}{3}\binom{13}{1}}{\binom{52}{13}}.$$

7. 等待时間的例子

在这一节里,为了考虑一些新型的样本空間,我們将避开組合分析的直接道路, 把占位問題稍許变化一下就可以得到这样新型的样本空間. 我們再一次地考虑把球随机地放入 n 个盒这个概念化了的"試驗". 然而,这时我們事先并不固定球的个数 r, 而是让球一个接一个地投入小盒,一直到我們指定的結果出現为止. 我

們将要仔細考虑下面两种可能的結果：（i）球一个接一个地往盒里随机地放，一直到有一个球放入某一个已經有球的盒为止．这个过程一直到有一个盒已經放入一个球时才結束．（ii）我們固定一个盒（譬如說，第一个盒），当这固定的盒还空时，我們就随机地放小球入盒，也就是說当这个盒一旦放入球时，这个过程就停止．

只要把这个模型稍許解释一下就会說明問題．

例．（a）生日．在例（3.d）中，一年中 $n=365$ 天对应为 365 个盒，人对应为球．我們的模型（i）現在可以这样說：如果我們一个接一个地随机地选人，問需要抽多少次才能发現有两个人同一天生日？模型（ii）輪到我的生日在样本中出現时需要抽多少人．

（b）鑰匙問題．一个人想打开他的門．他有 n 把鑰匙，但是只有一把能打开他的門．由于他事先不知哪一把能打开門，他随机地用这些鑰匙去試开，因此，每一把鑰匙在每一次試驗中被抽到的概率都是 n^{-1}，而且含有同样多的試驗次数的全部可能結果都是相似的．这个人恰巧在第 r 次試驗把門打开的概率是多少？这是模型（ii）的一个特别情形．把这种作法与問題（10.11）的更系統的探討比較一下是很有趣的；也可参看 V，問題 5．

（c）在前面的例子中，我們可以用从任意的总体中抽样来代替从鑰匙中抽样，譬如說，我們可以用收集贈劵来代替抽取鑰匙．我們再問：我們可望什么时候出現第一次重复，我們指望的那一张贈劵第一次出現是什么时候．

（d）錢币和骰子．在第一章例（5.a）中，一个錢币一直扔到出現正面为止．这是模型（ii）当 $n=2$ 的一个特殊情况．擲一顆骰子，一直到第一次出現么点为止，这也是模型（ii）当 $n=6$ 的一个特例．（在第 10 节問題 21，22，36；第 11 节問題 12 还考虑了其它一些等待时間．）

我們首先討論概念化的簡单的模型（i）．为了方便起見，我們利用符号 $(j_1, j_2, \cdots, j_r)$ 表示第 1，第 2，$\cdots$，第 r 个球放入第 j_1，j_2，$\cdots$，j_r 个盒，而且过程在第 r 次試驗結束．这就意味着 j_i 都是

1到n之間的整数；此外 $j_1, \cdots, j_{r-1}$ 都是不相同的，但 j_r 等于 $j_1, \cdots, j_{r-1}$ 中的某一个．每一个这样的排列都代表一个样本点．因为第一次出現某一个盒已經放了球不可能发生在第二次放球以前，也不可能发生在第 $n+1$ 次放球以后，所以，r 只可能取 $2, 3, \cdots, n+1$ 中的某一个值．現在的問題和过去老的模型(把固定数目的球放入 n 个盒中）之間的联系使得我們对每一个恰巧包含 r 个球的样本点賦以概率 n^{-r}．我們将要証明：这种作法是允許的(即是这些概率加起来等于1)，而且它可以推出許多合理的結果．

对于一个固定的 r，全部形如$(j_1, \cdots, j_r)$的样本点的总和代表事件"过程在第 r 步結束"．数 $j_1, \cdots, j_{r-1}$ 可以有 $(n)_{r-1}$ 种不同的选取法；而 j_r 只能取 $j_1, \cdots, j_{r-1}$ 这 $r-1$ 个数．由此推出，过程在第 r 步結束的概率为

$$q_r = \frac{(n)_{r-1} \cdot (r-1)}{n^r} = \\ = \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{r-2}{n}\right) \cdot \frac{r-1}{n}, \qquad (7.1)$$

其中 $q_1 = 0$，$q_2 = \frac{1}{n}$．过程在第 r 步以后結束的概率 $p_r = 1 - (q_1 + q_2 + \cdots + q_r)$，或者 $p_1 = 1$，

$$p_r = \frac{(n)_r}{n^r} = \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{r-1}{n}\right), \qquad (7.2)$$

上式可以用归納法得出．特別地，$p_{n+1}=0$，和 $q_1 + \cdots + q_{n+1} = 1$ 是显然的．此外，当 $n = 365$ 时，公式(7.2)推出(3.2)，一般，我們的新模型可以得出前面包含固定个数的球的模型的同样的結果．

模型(ii)与模型(i)之不同点在于模型(ii)要依賴一个无穷的样本空間．序列$(j_1, \cdots, j_r)$現在都滿足下述条件：$j_1, \cdots, j_{r-1}$ 都不等于給定的数 $a \leqslant n$，而 $j_r = a$．此外，沒有充分的理由說明过程为什么将要結束．对于一个固定的 r，我們可对每一个形如$(j_1, \cdots, j_r)$的样本点，都賦以概率 n^{-r}．$j_1, \cdots, j_{r-1}$ 中每一个都有 $n-1$ 种不同的选法，而 j_r 的选法只有一种．所以过程在第 r

步結束的概率为

$$q_r^* = \left(\frac{n-1}{n}\right)^{r-1} \cdot \frac{1}{n}, \quad r = 1, 2, \cdots. \tag{7.3}$$

把这个几何級数加起来，我們发現 $q_1 + q_2 + \cdots = 1$. 因此，这些概率之和为 1，所以，我們沒有必要引进一个样本点，它代表从来沒有一个球放入給定的盒 a. 过程在第 r 步以后才結束的概率

$$p_r^* = 1 - (q_1^* + \cdots + q_r^*),$$

因此，我們有

$$p_r^* = \left(1 - \frac{1}{n}\right)^r, \quad r = 1, 2, \cdots, \tag{7.4}$$

这是我們所期望的.

分布$\{p_r\}$和$\{p_r^*\}$的中位数我們定义为这样的 r：它使 p_r, p_r^* 在 $\frac{1}{2}$ 附近，大致上說就是这样，过程在中位数以后还継續的概率和在中位数以前結束的概率差不多.(在例(3.d)中，中位数为$r=23$.) 为了計算中位数，正如我們在(3.4)中所作的一样，我們要取对数. 当 r 相对 n 来說較小时，我們看出$-\log p_r$ 差不多等于 $r^2/2n$. 这就推出$\{p_r\}$的中位数大致上等于$(n \cdot 2 \cdot \log 2)^{\frac{1}{2}}$，或者近似地为 $\frac{6}{5} n^{\frac{1}{2}}$. 有趣的是，当总体的大小的平方根增加时，其中位数也增加. 为了比較，$\{p_r^*\}$的中位数大約为 $n \cdot \log 2$，或 $0.7n$，它对 n 是綫性地增加. 在模型(ii)中，等待时間超过 n 的概率为$(1 - n^{-1})^n$，或者近似地为 $e^{-1} = 0.36788\cdots$.

8. 二項式系数

当 n 是一个正整数时，我們曾經使用过二項式系数$\binom{n}{r}$，但是，可以很方便地把它的定义推广. 方程(2.1)中所引进的数$(x)_r$:

$$(x)_r = x(x-1)\cdots(x-r+1) \tag{8.1}$$

当 r 是一个正整数时对一切实数 x 都有定义. 当 $r=0$ 时，我們令 $(x)_0 = 1$. 于是

$$\binom{x}{r} = \frac{(x)_r}{r!} = \frac{x(x-1)\cdots(x-r+1)}{r!} \tag{8.2}$$

对全部实数 x 和全部正整数 r 定义为二项式系数．当 $r=0$ 时，如(4.4)一样，令 $\binom{x}{0}=1$ 和 $0!=1$．对于負整数 r，我們定义

$$\binom{x}{r} = 0 \quad (r<0). \tag{8.3}$$

如果 r 不是整数时，我們不用 $\binom{x}{r}$ 这一符号．

利用这个定义，我們很容易地驗証

$$\binom{-1}{r} = (-1)^r, \binom{-2}{r} = (-1)^r(r+1). \tag{8.4}$$

下面将要用到三个很重要的性質． 第一，对于每一个正整数 n 来說

$$\binom{n}{r} = 0 \quad \text{当 } r>n \text{ 或 } r<0. \tag{8.5}$$

第二，对每一个 x 和每一个正数 r 来說

$$\binom{x}{r-1} + \binom{x}{r} = \binom{x+1}{r}. \tag{8.6}$$

这些关系都可以由定义直接驗証．下一个关系的証明在一些微积分教科书中可以找到：对于任何数 a 和一切 $-1<t<1$，我們都有牛頓二項式公式

$$(1+t)^a = 1 + \binom{a}{1}t + \binom{a}{2}t^2 + \binom{a}{3}t^3 + \cdots. \tag{8.7}$$

如果 a 是一个正整数，右端諸項中次数高于 t^a 的項都为 0，从而公式对一切 t 都成立．如果 a 不是正整数，右端是一个无穷級数．

应用(8.4)，我們发現当 $a=-1$ 时，展式(8.7)化为几何級数

$$\frac{1}{1+t} = 1 - t + t^2 - t^3 + t^4 - + \cdots, \quad -1<t<1. \tag{8.8}$$

积分(8.8)，我們得到另外一个公式，它在以后是很有用的，这个公式就是自然对数的泰勒展式

$$\log(1+t) = t - \frac{1}{2}t^2 + \frac{1}{3}t^3 - \frac{1}{4}t^4 + \cdots,$$

$$-1<t<1. \tag{8.9}$$

(8.9)的两个另外的形式更常用．以$-t$代t，我們得到

$$\log\frac{1}{1-t}=t+\frac{1}{2}t^2+\frac{1}{3}t^3+\frac{1}{4}t^4+\cdots,$$

$$-1<t<1. \tag{8.10}$$

把最后这两个公式加起来我們得到

$$\frac{1}{2}\log\frac{1+t}{1-t}=t+\frac{1}{3}t^3+\frac{1}{5}t^5+\cdots,$$

$$-1<t<1. \tag{8.11}$$

当 $0<t<1$ 时，(8.10) 右端的数大于t但是小于 $t+t^2+t^3+\cdots=t/(1-t)$．因此，我們有双重不等式

$$e^{-\frac{t}{1-t}}<1-t<e^{-t},\quad 0<t<1. \tag{8.12}$$

在第 12 节中，将要从(8.7)推出許多有用的关系和恆等式．这里我們只注意n是正整数的情形，这时令 $t=1$，我們发現

$$\binom{n}{0}+\binom{n}{1}+\binom{n}{2}+\cdots+\binom{n}{n}=2^n. \tag{8.13}$$

这个公式有一个簡单的組合解释：把一个具有n个元素的总体分为两羣，第一羣的元素个数可以为 $k=0,1,\cdots,n$ 时，其分法的数目等于这个公式的左边． 另一方面，这样一种分法可以直接考虑每一个元素或者属于第一羣或者属于第二羣．（类似的推理証明：多項式系数(4.7)相加等于k^n.）

9. 史特令公式

分析概率論的一个重要工具蘊含在下述著名的定理[15]中：

史特令公式:

$$n!\sim(2\pi)^{\frac{1}{2}}n^{n+\frac{1}{2}}e^{-n}, \tag{9.1}$$

其中符号～表示两边之比当 $n\to\infty$ 时趋于 1.

这个公式是非常有用的，因为它有很大的理論价值，并且通过它可以得出一些精确的数值估計． 确实，(9.1)两边之差可以超过任何数，但是其相对誤差却是很小的．它下降得很迅速，甚至当n

很小时，史特令逼近都很精确． 事实上，(9.1)的右边以 0.9221 来逼近 1!，以 1.919 来逼近 2!，以 118.019 来逼近 5! = 120．其相对誤差分别为 8%，4%，2%． 对 10! = 3,628,800 来說，其近似值为 3,598,600，其相对誤差为 0.8%．对 100! 来說，其相对誤差只有 0.08%．

史特令公式的証明． 我們考虑

$$a_n = \log 2 + \log 3 + \cdots + \log(n-1) + \frac{1}{2}\log n \qquad (9.2)$$

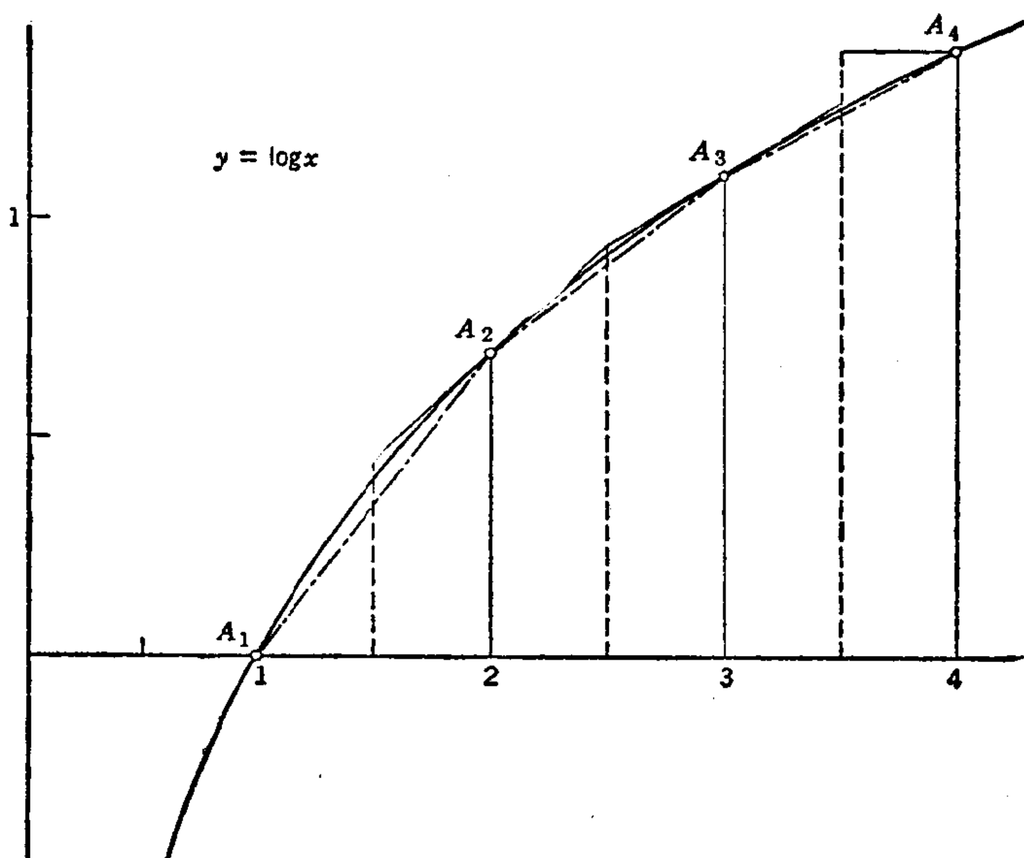


图 1． 說明史特令公式的导出，并更一般地用积分来說明逼近和．

它与 $\log n!$ 之差只有最后一項的 $\frac{1}{2}$ 倍．我們将要証明 a_n 表示二个不同的多边形的面积的差，而且由这一点給出 $\log n!$ 的两个界．图 1 說明 $n = 4$ 这一特殊情形的情况．写成下述形式：

$$a_n = \frac{1}{2}\{\log 1 + \log 2\} + \frac{1}{2}\{\log 2 + \log 3\} + \cdots + \\ + \frac{1}{2}\{\log(n-1) + \log n\} \qquad (9.3)$$

后，则易看出，a_n 等于端点为曲綫 $y=\log x$ 上的横坐标分别为 $1, 2, \cdots, n$ 的点 $A_1, A_2, \cdots, A_n$ 和 x 軸上的点 $(n, 0)$ 所构成的梯形的面积. 这个梯形在曲綫的內面，因此，它的面积小于由这条曲綫，x 軸和 $x=n$ 这条直綫所围成的区域的面积.

另一方面，$\log k$ 等于底边为 $k-\frac{1}{2}<x<k+\frac{1}{2}$，且界于曲綫在点 $A_k=(k, \log k)$ 的切綫下的梯形的面积. 由此推出，$\log(n-1)!$ 大于由 $y=\log x$，x 軸和垂直綫 $x=\frac{3}{2}$ 和 $x=n-\frac{1}{2}$ 所围成的区域的面积. 因为，显然 $\frac{1}{2}\log n$ 大于条形区域 $n-\frac{1}{2}<x<n$ 在曲綫下的面积，所以 a_n 大于曲綫，$x=\frac{3}{2}$，$x=n$ 和 x 軸所围成的面积. 換句話說，我們証明了

$$\int_{\frac{3}{2}}^{n}\log x\ dx<a_n<\int_1^n \log x\ dx. \tag{9.4}$$

$\log x$ 的不定积分为 $x\log x-x$，因此方程(9.4)化为二重不等式

$$\left(n+\frac{1}{2}\right)\log n-n+\frac{3}{2}\left(1-\log\frac{3}{2}\right)<\log n!<$$
$$<\left(n+\frac{1}{2}\right)\log n-n+1. \tag{9.5}$$

为了簡省起見，令

$$\delta_n=\log n!-\left(n+\frac{1}{2}\right)\log n+n. \tag{9.6}$$

于是 $1-\delta_n$ 是(9.5)最右边的数和 $\log n!$ 之差，也就是說，$1-\delta_n$ 等于曲綫 $y=\log x$ 与多边形 $A_1A_2\cdots A_n$ 之間的面积. 由此推出，δ_n 是单調下降的. 但是，由(9.5)我們有 $\frac{3}{2}\left(1-\log\frac{3}{2}\right)<\delta_n<1$. 所以我們断言，$\delta_n$ 趋于一个界于 1 和 $\frac{3}{2}\left(1-\log\frac{3}{2}\right)$ 的极限. 把这个极限用 $\log c$ 来記之，則我們有

$$\delta_n\to\log c,\ \text{其中}\ 2.45<c<2.72. \tag{9.7}$$

用对数的符号来写的話，史特令公式化为(9.7)，其中 $c=(2\pi)^{\frac{1}{2}}$(或者近似地为 2.507)．因为 π 可以用很多不同的方法定义，为了我們的目的，最簡单和最自然的是定义 $\pi=c^2/2$．利用这个定义，我們就得到史特令公式，但是，还留下一点需要証明，那就是这样定义的常数是否与 π 的其它公式一致．这个事实留在第七章中来証明，因此，我們完成了史特令公式的証明．

精确化． 史特令公式可以通过增加一項来进一步精确化．虽然我們并不需要应用如此精确的公式，但是，我們仍然在这里給出下列的二重不等式[16]的証明

$$(2\pi)^{\frac{1}{2}}n^{n+\frac{1}{2}}e^{-n+1/(12n+1)}<n!<(2\pi)^{\frac{1}{2}}n^{n+\frac{1}{2}}e^{-n+1/12n}. \tag{9.8}$$

为了証明(9.8)，我們注意

$$\delta_n-\delta_{n+1}=\left(n+\frac{1}{2}\right)\log\frac{n+1}{n}-1=$$

$$=\frac{1}{3(2n+1)^2}+\frac{1}{5(2n+1)^4}+\cdots \tag{9.9}$$

(最后一个式子在(8.11)中令 $t=1/(2n+1)$ 即可推出)．我們把(9.9)最右边的和的各項系数 $\frac{1}{5}$，$\frac{1}{7}$，$\frac{1}{9}$，… 都增大为 $\frac{1}{3}$，这就变为公比为 $(2n+1)^{-2}$ 的几何級数，因此

$$\delta_n-\delta_{n+1}<\frac{1}{3[(2n+1)^2-1]}=\frac{1}{12n}-\frac{1}{12(n+1)}. \tag{9.10}$$

所以，$\delta_n-1/12n$ 单调上升．因为这个序列的极限由史特令公式所給出，取逆对数，我們就得到(9.8)的第二个不等式．第一个不等式可以由(9.9)用类似的办法証明，这时，只需注意

$$\delta_n-\delta_{n+1}>\frac{1}{3(2n+1)^2}>\frac{1}{12n+1}-\frac{1}{12(n+1)+1}. \tag{9.11}$$

逼近式(9.8)是很精确的，甚至当 $n=1$ 时，这个公式所給出的两个界都是 0.9958…和1.0023…．(9.8)中給出的上界还稍为好一点(参看(12.28))．当 $n=2$ 时，它給出 2.0007；当 $n=5$ 时，我們得到 120.01…；当 $n=10$ 时，其前 5 位有效数字都正确．

問　　題

注意：第 11，12 两节包含着不同性质的問題以及課文的各种补充．

10. 練习和例子

注意：在每一种情形都假定所有的排列具有相同的概率.

1.若每一个人的名字[1]：(a)恰巧有 2 个字；(b)至多有 2 个字；(c)至多有 3 个字；試問姓名不相同的人共有多少？

2.不同顏色的 2 个城形棋子(国际象棋)，在棋盘上能彼此吃掉的放法有多少？

3.电碼中，普通的字是以“长划”和“点”二种記号容許重复应用来表示的. 問不超过 10 个記号可以表示的字有多少？

4.每块骨牌刻上 2 个数，骨牌是对称的，所以同一块牌上的数对是沒有次序的. 如果用 $1, 2, \cdots, n$ 等数，問能刻出多少不同的牌来？

5.数 $1, 2, \cdots, n$ 依任意順序排列，試求数列中：(a)1 与 2；(b) 1, 2 和 3，以(a)，(b)为序相継出現的概率.

6.(a)在三个随机数碼中，一个数碼重复发生 2 次，1 次或 0 次的概率为何？(b)对四个随机数碼的情形，把上述問題再作一遍.

7.試求在 r 个随机数碼的样本中，沒有二个数碼相同的概率 p_r，并用史特令公式，估計 p_{10} 的数值.

8.在 k 个随机数碼中：(a)0 不出現；(b)1 不出現；(c)0 与 1 都不出現；(d)0和1二数至少有一个不出现的概率为何？令 A 和 B 分別表示事件(a)和(b)，試用 A 和 B 把其余的事件也表示出来.

9.n 个球随机放入 n 个盒中，問恰有一盒空着的概率为何？

10.一个停車場有 12 个位置列成一行. 某人发現有 8 个位置停了車，而有 4 个接連(构成一个連貫)的位置空着. 这种发現令人惊奇(是非随机性的表示)嗎？

11.一人有 n 个鑰匙，其中只有一把能开开他的門. 他逐个地用它們去試开(抽样是无放回的)这可能要 1 次，2 次，…，n 次才能把門开开. 証明这 n 种不同的結果的概率都是$1/n$.

12.假定 n 支手杖的每一支都折断成一长一短的二个小段，把这 $2n$ 个小段任意排成 n 对，組成新手杖. 試求：(a)这些小段都連接成原来的样子的概率；(b)所有长的小段都与短的小段配对的概率[2].

---

1) 以 26 个字母排列成人的名字——譯者注.

2) 当細胞被有害的放射綫照射的时候，一些染色体就会破裂，它們的作用正如此习題中的“手杖”一样. “长的”那一边就是含有所謂中心体的那一边. 如果两个“长”的部分相結合或两个“短”的部分相結合，那么此細胞就死亡. 參看[17].

13.統計假設的检驗． 某工作人員在某一个星期里，曾經接見訪問 12 次，所有这 12 次的訪問恰巧都是在星期二或星期四． 試求該事件的概率．(是否可以断定他只在星期二和星期四接見訪問者?)

14(續前)．这 12 次訪問沒有一次是星期日，是否可以断言星期日他根本不会客?

15.一个匣子里，有 90 只好的螺絲釘，10 只坏的螺絲釘．如果从中任意取用 10 只螺絲釘，恰巧都是好的螺絲釘的概率为何?

16.从 5 个記号 a, b, c, d, e 的总体中，抽取大小为 25 的一个样本，求样本中恰巧含有每类記号 5 个的概率．并把 a 等同于数字 0 和 1，b 等同于数字 2 和 3，等等，用随机数的表来校核这个結果[1]．

17.若 n 个人站成一个横列，其中有 A 和 B 二人，間夹在 A 和 B 之間，恰有 r 个人的概率为何? 如果他們不是站成一列而是站成一圈，試証这个概率与 r 无关，而且它就是 $\frac{1}{n-1}$．

18.把 3 顆骰子擲二次，每次出現点子的花样是相同的的概率为何? 如果：(a)骰子可分辨的；(b)不可分辨的，試分别求之．

19.試証，4 顆骰子擲一次至少出現一个"么"点的可能性比 2 顆骰子擲 24 次至少出現一个双"么"点的可能性为大．(这个問題回答了所謂台·曼来的詩論．台·曼来是一个賭徒，他認为这二个概率应該是相等的，由于賭輸他曾譴責过数学．)

20.从一个由 n 个元素构成的总体中抽出一个大小为 r 的样本．求指定的 N 个元素不包含在样本中的概率，假定(a)无放回；(b)有放回的．当 (i) $n=100, r=N=3$；(ii) $n=100, r=N=10$ 时，比較上述两种抽样法所得到的概率．

21.謠言的传播． 在一个拥有 $n+1$ 个居民的城市里，某一个人告訴第二个人一个謠言，而第二人又把謠言告訴第三个人，如此等等．在每一步中，謠言的接收者都是随机地从 n 个居民中挑选的． 求下述两事件的概率：謠言传播了 r 次后，(a)还沒有回到第一个造謠者，(b)沒有一个人两次地听到謠言．当每一次都把謠言同时告訴由城市中随机选取的 N 个居民时，問上面两事件的概率等于多少．(最初的問題是 $N=1$ 的特别情形．)

22.书信的連續传遞． 在一个具有 $n+1$ 个人的集体里，有一个人——"祖先"，发两封信給两个人——"第一輩后代"．而这两个人又各发两封信給

1) 它們有时候是非常的相近，参看[18]．

别人，一般地，第 r 輩后代中的某一个人都随机地发两封信給别人。求出第1，2，…，r 輩后代都不包含祖先的概率。假定 n 充分大，求出分布的中位数。

23. 一个著名的問題。某家有四个女孩，她們去洗食具。在打破的四个食具中有三个是最小的女孩打破的，因此人家說她笨拙。她是否有理由申辯这完全是碰巧？討論这一題和球随机地放入盒中的联系。

24. 求出下述两事件的概率：(a) 12 个人的生日在 12 个不同的月份（假定任何一个人生于 12 个月中之任一月都是等概的）；(b) 6 个人的生日恰巧在两个月中。

25. 給定 30 个人，求出 12 个月中有六个月恰巧包含两个人的生日；有六个月恰巧包含三个人的生日的概率。

26. 一个房間里有 n 双不同型号的鞋子。今从其中随机地抽取 $2r$ 只 $(2r<n)$，求下面三个事件的概率：(a) 沒有一双同型号的；(b) 恰有一双同型号的；(c) 恰有两双同型号的。

27. 一輛車子停在有 N 輛車子的行中間（不在两端）。当他过一会再回来的时候，发現 N 个位置中恰巧仍有 r 个位有車。問两个相邻的位置是空的概率为何？

28. 把拥有 $2N$ 个男孩和 $2N$ 个女孩的一羣孩子分为两羣，每羣各 $2N$ 个孩子，求每一羣中男女数目相等的概率 p。并用史特令公式估計 p。

29. 証明：在打桥牌中，坐在西方的人恰巧拿住 k 个爱司的概率 p 与任取 13 张牌其中恰有 k 个爱司的概率相等。（直观上这是很显然的。不过要注意，这两个概率所考慮的試驗是不同的，因为在后一种情形中，13 张牌是任意抽取的，而在前一种情形中，52 张牌都要分配完的。）

30. 在桥牌游戏中，东家和南家各有 m，n 张黑桃的概率和从整付桥牌中随机地取两付（每付 13 张），第一付有 m 张黑桃第二付有 n 张黑桃的概率同。

31. 南北两家共有 k 个爱司 $(k=0, 1, 2, 3, 4)$ 的概率为何？

32. 令 a, b, c, d 为满足 $a+b+c+d=13$ 的四个非負整数，在一次桥牌游戏中，求东南西北各家分别拿到 a, b, c, d 张黑桃的概率 $p(a, b, c, d)$。描述一个把紅球黑球放入盒中的模型，而把这个問題作为一个特例。

33. 利用 32 題的結果，(a) 当 $a=5, b=4, c=3, d=1$ 时；(b) $a=b=c=4, d=1$ 时；(c) $a=b=4, c=3, d=2$ 时，求出 $p(a, b, c, d)$ 之值。

注意，这三种情况有本质的区别。

34. 令 a, b, c, d 为满足 $a+b+c+d=13$ 的整数，求出在一付桥

牌中含有 a 个黑桃 b 个紅心 c 个方块和 d 个梅花的概率 $q(a, b, c, d)$，并証明这个問題不可能化为一个把 13 个球随机地放入 4 个盒中的模型. 为什么?

35. r 张桥牌中爱司的分布. 計算随机抽取 r 张桥牌，其中恰有 0，1，…，4 张爱司的概率 $p_0(r)$，$p_1(r)$，…，$p_4(r)$. 驗証 $p_0(r)=p_4(52-r)$.

36（續上）. 等待时間. 如果取出一张一张牌，求出第 1，…，第 4 张爱司在第 r 次抽取中出現的概率 $f_1(r)$，…，$f_4(r)$. 猜測第 1，…，第 4 张爱司的等待时間的中位数，并从而算出它們来.

37. 如果两付桥牌中每付都有 r 张，而且(a)它們都是从同一付桥牌中取出的；(b)它們从两付桥牌中取出的，試求每一付牌中恰有 k 个爱司的概率. 証明：当 $r=13$ 时，問題(a)的概率与指定两个桥牌游戏者各获 k 个爱司的概率一样.

38. 印錯. 每一頁书中都有 N 个印刷符号可能誤印. 全书共有 $n=500$ 頁，$r=50$ 个印錯的符号. 証明(a)第 1，第 2，…，第 n 頁分别含有 r_1，r_2，…，r_n 个錯的符号的概率等于

$$\binom{N}{r_1}\binom{N}{r_2}\cdots\binom{N}{r_n}\div\binom{nN}{r};$$

(b)当 N 充分大后，这个概率可以用(5.5)来逼近，并推出 r 个錯誤分布在 n 頁的問題近似地与 r 个球随机地放入 n 个盒的問題相吻合.（注意：这个可以作为弗米-迪拉克統計的一个普遍的极限性質. 参看第 5 节.）

注意：下面的問題与第 5 节的材料有联系.

39. 若把 r_1 个不可分辨的第一类东西和 r_2 个不可分辨的第二类东西放入 n 个盒中，試求可区分的排列总数.

40. 把 r_1 个骰子和 r_2 个小錢一起扔一次，試問有多少可以区分的結果.

41. 把 r_1 个白球，r_2 个黑球，r_3 个紅球进行排列，試問有多少种可以区分的排列方式.

42. 把 52 张桥牌随机排列，試問沒有两张爱司紧邻的概率为何?

43. 电梯. 在例(3.c)中，在开始时載有 7 位乘客的一架电梯在有 10 层楼房的每一层上停留，这 7 个乘客出电梯有各种各样可能的排列，例如我們用(3, 2, 2)表示有三个乘客在某一层同时出去，另外又有两个乘客在另一层出去，最后两个乘客在某一层一同出去. 这种可能的排列从(7)到(1,1,1,1,1,1,1)共有 15 种，試求出对应于这 15 种排列的15个概率.

44. 生日. 求出 22 个人的生日的各种不同形式的概率.

45. 如果一付扑克恰巧是下列事件之一，試分別求其概率：

(a) 最大同花("10", "J", "Q", "K", "A"为同一花色)；

(b) 四同点(4张牌的面值相同)；

(c) 滿堂(有一对且有三张同面值)；

(d) 順子(5张牌的順序連續，但不考慮花色)；

(e) 三同点(3张的面值相同与2张不同的牌)；

(f) 两对(4张牌成两对与另外的一张牌)；

(g) 一对(两张牌同面值与另外三张不同的牌).

11. 問題和理論性的附录

1. 一个总体中有 n 个元素，其中有 np 个紅的 nq 个黑的($p+q=1$). 有放回地抽取大小为 r 的随机样本. 試証其中恰有 k 个紅的的概率为

$$\binom{r}{k} p^k q^{r-k}. \tag{11.1}$$

2. 一个超几何分布的极限定理. 如果 n 充分大，而且 $n_1/n=p$，則由(6.1)和(6.2)所給出的概率 q_k 与(11.1)很近似. 更精确地說有

$$\binom{r}{k}\left(p-\frac{k}{n}\right)^k\left(q-\frac{r-k}{n}\right)^{r-k}<q_k<\binom{r}{k} p^k q^{r-k}\left(1-\frac{r}{n}\right)^{-r}. \tag{11.2}$$

把它和前一个問題比較，則表明：对大的总体来說，有放回的抽样和无放回的抽样在实际上并无什么区别.

3. 从一个拥有 n 个元素的总体中无放回地抽取一个大小为 r 的随机样本. 某指定的 N 个元素全部被抽中的概率 u_r 为

$$u_r=\binom{n-N}{r-N}\div\binom{n}{r}. \tag{11.3}$$

(对有放回的抽样来說，其对应的公式由(11.10)所給出，它不可能由此直接推出. (11.3)的另外一种形式，請参看第四章問題9.)

4. 极限形式. 如果 $n\longrightarrow\infty$ 和 $r\longrightarrow\infty$ 滿足 $r/n\longrightarrow p$，則 $u_r\longrightarrow p^N$(参看問題13).

注意：問題5—問題13与著名的占位問題有联系(馬克司威尔-波茨曼統計)：也就是說，r 个球分布在 n 个盒中，这 n^r 种可能中每一种可能的概率都是 n^{-r}[1].

---

1) 問題5—19在量子統計中，在攝影感光版的理論中，在G-M計数器等等中都有应用. 因此，在物理学文献中这些公式都是經常討論經常出現的，通常是沒有把它們的重要的和本質的初等性質加以抽象. 几乎所有的在这章开始时所引进的問題在惠瓦茨(Whitworth)的书中都碰得到(虽然形式稍有不同).

5. 某一个指定的盒中恰有 k 个球的概率 p_k 由二項分布 (4.5) 所給出. 最大可能的整数 ν 满足 $(r-n+1)/n<\nu\leqslant(r+1)/n$. (换句話說,它說明 $p_0<p_1<\cdots<p_{\nu-1}\leqslant p_\nu>p_{\nu+1}>\cdots>p_r$; 参看問題 15.)

6. 极限形式. 如果 $n\longrightarrow\infty$ 和 $r\longrightarrow\infty$ 使得每个盒中球的平均数 $\lambda=r/n$ 保持常数,則

$$p_k\longrightarrow \frac{e^{-\lambda}\lambda^k}{k!}. \tag{11.4}$$

这就是普阿松分布,它将在第四章中討論;参看問題 16.

7. 令 $A(r, n)$ 为 n 个盒中沒有一个是空的可能分布的个数. 用組合分析的推理証明

$$A(r, n+1)=\sum_{k=1}^{r}\binom{r}{k}A(r-k, n), \tag{11.5}$$

并推出

$$A(r, n)=\sum_{\nu=0}^{n}(-1)^\nu\binom{n}{\nu}(n-\nu)^r. \tag{11.6}$$

提示: 应用归納法;假定(11.6)成立,并且以此表示(11.5)中的 $A(r-k, n)$. 交换求和的次序,并用二項公式去表 $A(r, n+1)$ 为二个单重和的差. 代第二个和中的 $\nu+1$ 以一个新的求和指标并应用(8.6).

注意: 公式(11.6)是一个古典問題的理論上的結果,用它来直接計算概率是不方便的,例如,某一乡村有 1900 个居民,用这个公式来算一年中每一天都有这 1900 个居民的生日的概率 x 就很麻煩. 在第四章第 2 节中,我們将用另一种办法推出公式(11.6),并且得到一个簡单的近似公式 (对上面的例子来說, x 近似地为 0.135).

8. 証明恰有 m 个盒是空的可能分布的种数为

$$E_m(r, n)=\binom{n}{m}A(r, n-m)=\binom{n}{m}\sum_{\nu=0}^{n-m}(-1)^\nu\binom{n-m}{\nu}(n-m-\nu)^r. \tag{11.7}$$

9. 不用上面的結果来直接証明恰有 m 个盒是空的概率 $p_m(r, n)=n^{-r}E_m(r, n)$ 满足

$$p_m(r+1, n)=p_m(r, n)\frac{n-m}{n}+p_{m+1}(r, n)\frac{m-1}{n}. \tag{11.8}$$

10. 应用問題 7 和 8 的結果,用直接計算的方法来証明(11.8)成立. 証明这种方法可以給出(11.6)一个新的推导(对 r 作归納法).

11. 从(11.6)和問題 8 推出: 大于等于 m 个盒是空的概率为

$$\binom{n}{m}\sum_{\nu=0}^{n-m}(-1)^\nu\binom{n-m}{\nu}\left(1-\frac{m+\nu}{n}\right)^r\frac{m}{m+\nu}. \tag{11.9}$$

(当 $m\geqslant n$ 时，这个式子为 0，这是意料中的事情.)

12. N 个指定的盒中，每一个都被占有的概率为

$$u(r,n)=n^{-r}\sum_{k=0}^{r}\binom{r}{k}A(k,N)(n-N)^{r-k}. \tag{11.10}$$

由此断言

$$u(r,n)=\sum_{\nu=0}^{r}(-1)^\nu\binom{N}{\nu}\left(1-\frac{\nu}{n}\right)^r. \tag{11.11}$$

(应用二项定理. 对 $N=n$ 我們有 $u(r,n)=n^{-r}A(r,n)$. 注意：(11.11) 类似于有放回的抽样[1] 的公式(11.3). 另一个推导参看第四章問題 8 .)

13. 极限形式. 如果换作問題 4 中的极限描述，我們有 $u(r,n)\longrightarrow(1-e^{-p})^N$.

注意：在問題 14—19 中，r 和 n 的意义与以前一样，但是我們假定球是不可辨別的，而且全部可辨别的排列都是等概的(波司-爱因斯坦統計).

14. 某一个指定的盒中恰有 k 个球的概率为

$$q_k=\binom{n+r-k-2}{r-k}\div\binom{n+r-1}{r}. \tag{11.12}$$

15. 証明：当 $n>2$ 时，任何一个确定的盒中球的个数的最大可能的值为 0，或者更精确地說，$q_0>q_1>\cdots$ (参看問題 5).

16. 极限定理. 令 $n\longrightarrow\infty$ 和 $r\longrightarrow\infty$ 使得每一个盒中平均的質点个数 r/n 趋于 λ. 于是

$$q_k\longrightarrow\frac{\lambda^k}{(1+\lambda)^{k+1}}. \tag{11.13}$$

(右边了解为几何分布.)

17. 恰有 m 个盒是空的的概率为

$$p_m=\binom{n}{m}\binom{r-1}{n-m-1}\div\binom{n+r-1}{r}. \tag{11.14}$$

---

1) 注意：$u(r,n)$ 可以解释为当第 N 个元素加入样本时所需之等待时間小于 r 的概率. 其結果可以应用到随机抽取数字中去：其中 $u(r,10)-u(r-1,10)$ 为长为 r 的序列中包含了由全部 10 个数字所构成的全集的概率. 这个可以用来作随机检驗. 格林伍德(Greenwood)(参看[19])把这些分布列成一个表，并把它与一些眞实的計算进行比较，这些計算是对应于 π 的前 2035 位小数和 e 的前 2486 位小数的等待时間. 全部 10 个数字构成的全集的等待时間的中位数是 27. 这个等待时間超过 50 的概率大于 0.05，而等待时間超过 75 的概率大約为 0.0037.

18. 某指定的 m 个盒中的球的总数恰为 k 的概率为

$$q_j(m)=\binom{m+j-1}{m-1}\binom{n-m+r-j-1}{r-j}\div\binom{n+r-1}{r}. \tag{11.15}$$

19. **极限形式**. 如果换作問題 4 中的极限描述，我們有

$$q_j(m)\longrightarrow\binom{m+j-1}{m-1}\frac{p^j}{(1+p)^{m+j}}. \tag{11.16}$$

(右边是在第六章中引进的負二項分布的特殊情形.)

关于連貫的定理. 在問題 20—25 中，我們考虑由 r_1 个 α 和 r_2 个 β 所构成之排列，而且假定所有排列都是等可能的[参看例(4.d)]. 这一組問題与第 5a 节有联系.

20. 在排列中恰有 k 个連貫(不分种类)的概率当 $k=2\nu$ 是偶数时为

$$P_{2\nu}=2\binom{r_1-1}{\nu-1}\binom{r_2-1}{\nu-1}\div\binom{r_1+r_2}{r_1}, \tag{11.17}$$

当 $k=2\nu+1$ 为奇数时，其概率为

$$P_{2\nu+1}=\left\{\binom{r_1-1}{\nu}\binom{r_2-1}{\nu-1}+\binom{r_1-1}{\nu-1}\binom{r_2-1}{\nu}\right\}\div\binom{r_1+r_2}{r_1}. \tag{11.18}$$

21(續上). 推出連貫的个数的最大可能数值为滿足

$$\frac{2r_1r_2}{r_1+r_2}<k<\frac{2r_1r_2}{r_1+r_2}+3$$

的整数 k. (提示: 考虑比 $P_{2\nu+2}\div P_{2\nu}$ 和 $P_{2\nu+1}\div P_{2\nu-1}$.)

22. 排列中以一个长为 $\nu\geqslant 0$ 的 α-連貫开始的概率为 $(r_1)_\nu r_2\div(r_1+r_2)_{\nu+1}$. (提示: 选取 ν 个 α 和跟随其后的 β.) 当 $\nu=0$ 时，定理蘊含了什么呢?

23. 恰有 k 个 α-連貫的概率为

$$\pi_k=\binom{r_1-1}{k-1}\binom{r_2+1}{k}\div\binom{r_1+r_2}{r_1}. \tag{11.19}$$

提示: 这可以由第 5 节的引理的第二部分直接推出. 另外，方程(11.19)可以由(11.17)和(11.18)推出，不过手續比較繁难.

24. 第 n 个 α 恰在 m 个 β 之前的概率为

$$\binom{r_2-m+n-1}{r_2-m}\binom{n_1-n-1}{m}\div\binom{r_1+r_2}{r_1}. \tag{11.20}$$

25. 全部 α 分居于 k 个連貫中，其中有 k_1 个长度为 1，k_2 个长度为 2，…，k_ν 个长度为 $\nu(k_1+k_2+\cdots+k_\nu=k)$ 的概率为

$$\frac{k!}{k_1!k_2!\cdots k_\nu!}\binom{r_2+1}{k}\div\binom{r_1+r_2}{r_1}. \tag{11.21}$$

12. 二項式系数的一些問題和恆等式

1.对于整数 $n\geq 2$ 来說有

$$
\begin{aligned}
&1-\binom{n}{1}+\binom{n}{2}-+\cdots=0,\\
&\binom{n}{1}+2\binom{n}{2}+3\binom{n}{3}+\cdots=n2^{n-1},\\
&\binom{n}{1}-2\binom{n}{2}+3\binom{n}{3}-+\cdots=0,\\
&2\cdot 1\binom{n}{2}+3\cdot 2\binom{n}{3}+4\cdot 3\binom{n}{4}+\cdots=n(n-1)2^{n-2}.
\end{aligned}
\tag{12.1}
$$

(提示: 应用二項公式.)

2.証明: 对于正整数 n 和 k, 我們有

$$\binom{n}{0}\binom{n}{k}-\binom{n}{1}\binom{n-1}{k-1}+\binom{n}{2}\binom{n-2}{k-2}\cdots\pm\binom{n}{k}\binom{n-k}{0}=0. \tag{12.2}$$

更一般地[1],我們有

$$\sum\binom{n}{\nu}\binom{n-\nu}{k-\nu}t^{\nu}=\binom{n}{k}(1+t)^{k}. \tag{12.3}$$

3.对于任何的 $a>0$, 都有

$$\binom{-a}{k}=(-1)^{k}\binom{a+k-1}{k}. \tag{12.4}$$

如果 a 是一个整数,这个方程可以由对几何級数 $\Sigma x^{k}=(1-x)^{-1}$ 進行微商来証明.

4.証明

$$\binom{2n}{n}2^{-2n}=(-1)^{n}\binom{-\frac{1}{2}}{n}. \tag{12.5}$$

5.对于非負整数 n 和 r 和一切实数 a 来說,都有

$$\sum_{\nu=0}^{n}\binom{a-\nu}{r}=\binom{a+1}{r+1}-\binom{a-n}{r+1}. \tag{12.6}$$

(提示: 应用方程(8.5). $n=a$ 的特别情形是經常用到的.)

6.对于任意 a 和整数 $n\geq 0$ 来說,都有

$$\sum_{\nu=0}^{n}(-1)^{\nu}\binom{a}{\nu}=(-1)^{n}\binom{a-1}{n}. \tag{12.7}$$

(提示: 应用方程(8.5).)

---

1) 讀者回忆一下(8.5)的約定;如果 ν 跑遍全部整数, 則(12.3)中的和中只有有限多項非0.

7.对于正整数 r 和 k 来說，有

$$\sum_{\nu=0}^{r}\binom{\nu+k-1}{k-1}=\binom{r+k}{k}. \tag{12.8}$$

(a)用(8.6)証明这个等式. (b)証明(12.8)是(12.7)的特殊情形. (c)用归納法証明(12.8)給出第 5 节的引理的第一部分的一个新的証明.

8.在第 6 节中曾指出超几何分布諸項之和为 1. 这說明对任何正整数 a, b, n 来說都有

$$\binom{a}{0}\binom{b}{n}+\binom{a}{1}\binom{b}{n-1}+\cdots+\binom{a}{n}\binom{b}{0}=\binom{a+b}{n}. \tag{12.9}$$

用归納法証明之.(提示: 首先証明(12.9)对 $a=1$ 和全部 b 都成立.)

9(續上). 用比較方程

$$(1+t)^a(1+t)^b=(1+t)^{a+b} \tag{12.10}$$

两边 t^n 的系数的办法来更一般地証明: (12.9)对任意 a, b (及任意整数 n)都成立.

10.应用方程(12.9)証明

$$\binom{n}{0}^2+\binom{n}{1}^2+\binom{n}{2}^2+\cdots+\binom{n}{n}^2=\binom{2n}{n}. \tag{12.11}$$

11.应用方程(12.10)証明

$$\sum_{\nu=0}^{n}\frac{(2n)!}{(\nu!)^2(n-\nu)!^2}=\binom{2n}{n}^2. \tag{12.12}$$

12.証明: 对于整数 $0<a<b$ 有

$$\sum_{k=1}^{a}(-1)^{a-k}\binom{a}{k}\binom{b+k}{b+1}=\binom{b}{a-1}. \tag{12.13}$$

(提示: 应用 (12.14) 証明 (12.11) 是 (12.9) 的特别情形. 此外，再比較 $(1-t)^a(1-t)^{-b-2}=(1-t)^{a-b-2}$ 两边 t^{a-1} 的系数.)

13.从(12.9)用特殊方法推出恒等式

$$\binom{a}{k}-\binom{a}{k-1}+-\cdots\mp\binom{a}{1}\pm1=\binom{a-1}{k} \tag{12.14}$$

和

$$\sum_{\nu}(-1)^{\nu}\binom{a}{\nu}\binom{n-\nu}{r}=\binom{n-a}{n-r} \tag{12.15}$$

对正整数 k, n 和 r 都成立. [提示: 应用(12.14).]

14.应用方程(12.9)証明

$$\sum_{j=0}^{k}\binom{a+k-j-1}{k-j}\binom{b+j-1}{j}=\binom{a+b+k-1}{k}. \tag{12.16}$$

(提示：应用方程(12.4)的后部和前部.) 注意 $b=1, 2$ 的重要的特殊情形.

15.和第 11 节中的問題联系起来,注意：方程(11.12),(11.14),(11.15)和(11.16)都定义概率. 因此,在每一个方程中諸量之和为 1. 証明：这些关系分别地被(12.8),(12.9),(12.16)和二項定理所蘊含.

16.从第 11 节問題 7 中关于 $A(r, n)$ 的定义出发，可以推出：如果 $r<n$，則 $A(r, n)=0$，而且 $A(n, n)=n!$. 換句話說

$$\sum_{k=0}^{n}(-1)^{n-k}\binom{n}{k}k^r=\begin{cases}0 & \text{如果 } r<n,\\ n! & \text{如果 } r=n.\end{cases} \tag{12.17}$$

(a)用化 n 为 $n-1$ 的办法直接証明(12.17). (b)其次用考慮 $(1-e^t)^n$ 在 $t=0$ 的 r 次微商的方法来証明(12.17). (c) 从(11.11)出发而不从(11.6)出发来推广(12.17).

17.如果 $0\leqslant N\leqslant n$，用归納法証明：对每个整数 $r\geqslant 0$，都有

$$\sum_{\nu=0}^{N}(-1)^{\nu}\binom{N}{\nu}(n-\nu)_r=\binom{n-N}{r-N}r!. \tag{12.18}$$

(注意：当 $r<N$ 和 $r>n$ 时,右边为0.)用考虑 $t^{n-N}(t-1)^N$ 在 $t=1$ 的微商的办法驗証(12.18).

18.用归納法証明(应用二項定理)

$$\binom{n}{1}\frac{1}{1}-\binom{n}{2}\frac{1}{2}+\cdots+(-1)^{n-1}\binom{n}{n}\frac{1}{n}=$$

$$=1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}. \tag{12.19}$$

用积分恆等式 $\sum_{0}^{n-1}(1-t)^{\nu}=\{1-(1-t)^n\}t^{-1}$ 的办法来驗証(12.19).

19.証明对任何正整数 m 都有

$$(x+y+z)^m=\Sigma\frac{m!}{a!b!c!}x^a y^b z^c, \tag{12.20}$$

其中求和号是对全部滿足 $a+b+c=m$ 的非負整数 a, b, c 来求和的.

20.应用史特令公式証明

$$\binom{2n}{n}\sim(\pi n)^{-1/2}2^{2n}. \tag{12.21}$$

21.証明对任何正整数 a, b 都有

$$\frac{(a+1)(a+2)\cdots(a+n)}{(b+1)(b+2)\cdots(b+n)}\sim\frac{b!}{a!}n^{a-b}. \tag{12.22}$$

22.Γ-函数的定义如下：

$$\Gamma(x)=\int_0^\infty z^{x-1}e^{-z}dz, \tag{12.23}$$

其中 $x>0$。証明 $\Gamma(x)\sim(2\pi)^{1/2}e^{-x}x^{x-(1/2)}$。(注意: 如果 $x=n$ 是一个整数,則 $\Gamma(n)=(n-1)!$。)

23. 令 a 和 r 是任意正数,而 n 是一个正整数。証明

$$a(a+r)(a+2r)\cdots(a+nr)\sim Cr^{n+1}n^{n+(a/r)+1/2}e^{-n}. \tag{12.24}$$

[常数 C 等于 $(2\pi)^{1/2}/\Gamma(a/r)$。]

24. 应用前面一个問題的結果,証明

$$\frac{a(a+r)(a+2r)\cdots(a+nr)}{b(b+r)(b+2r)\cdots(b+nr)}\sim\frac{\Gamma(b/r)}{\Gamma(a/r)}n^{(a-b)/r}. \tag{12.25}$$

25. 証明下述史特令公式的另一形式:

$$n!\sim(2\pi)^{1/2}\left(n+\frac{1}{2}\right)^{n+1/2}e^{-(n+1/2)}. \tag{12.26}$$

26(續上). 应用課文里面的方法,証明

$$(2\pi)^{1/2}\left(n+\frac{1}{2}\right)^{n+1/2}e^{-(n+1/2)-1/24(n+1/2)}<n!<$$

$$<(2\pi)^{1/2}\left(n+\frac{1}{2}\right)^{n+1/2}e^{-(n+1/2)}. \tag{12.27}$$

27. 推广史特令公式,証明

$$n!\sim(2\pi)^{1/2}n^{n+1/2}\exp\left\{-n+\frac{1}{12n}-\frac{1}{360n^3}+\cdots\right\}. \tag{12.28}$$

*第三章[1] 扔錢币的起伏問題和随机徘徊

本章目的有二. 第一,通过它,将要說明极簡单的方法可以导出深刻而又重要的結果. 第二,在这一章中,我們将要第一次碰到一些理論上的結論,它不仅是我們所想象不到的,而且实在給直观和常識一个大大的振惊. 它們将会說明,一般人所接受的关于随机起伏的概念是沒有根据的,而且关于大数定律的含意在很大程度上被誤解[2].

这里所以插入这样的討論只因它有初等的性质,本书的主要論題在第五章再接續. 全书都不依賴这一章. 这一章里某些公式在后面討論初过和再現时将会重新出現,不过它們还要用分析方法重新推出. 方法的比較将是富有教益的而且是有趣的. 因此,讀者可按自己的判断閱讀本章或者与本书其余各章同时并进地閱讀,或者单独地閱讀. 为了对上述办法給以便利,这章书可以分两次閱讀: 主要的标題用通常的形式印出. 用小字排印的地方包含一些附加的內容(主要是与初过和再現現象有联系),因而初讀时可以略去. 第 7 节包含一个經驗的解释.

1. 一般討論

关于一般的随机起伏的相当丰富內容可以由下述的伯川(Bertrand)在 1887 年所发表的一个不显著的引理所推出,类似的

---

1) 这一章可以略去,或者和下面几章結合起来讀. 在第十章(大数定律),第十一章(初过时間),第十二章(再現事件),第十四章(随机徘徊)中都将要参考本章的內容. 不过,后面沒有明显地用到这些內容.

2) 虽然形式上我們只考慮扔錢币,但是其基本結論是有广泛的应用的. 事实上,安德生(E. Sparre Andersen)的惊人发現: 独立随机变量和的起伏理論的很多方面都具有純粹組合的性质,而且这样的随机变量中有很大一族都具有共同的性质. 特别言之两个反正弦律就是如此. 参看[20].

排列問題，在以选举問題命名的組合分析里曾吸引过学生們的兴趣[1]. 假定一次选举中,候选人 P 获得 p 张选票，候选人 Q 获得 q 张选票,其中 $p > q$. 在数票过程中經常是 P 所获之票比 Q 所获之票多的概率为$(p-q)/(p+q)$.

用数学的語言来說,这里我們所考慮的是由 $x = p + q$ 个符号 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_x$ 所构成的一些排列，$\varepsilon_1, \cdots, \varepsilon_x$ 中有 p 个正 1 (P 的选票) q 个負 1 (Q 的选票). 部分和 $s_k = \varepsilon_1+\varepsilon_2+\cdots+\varepsilon_k$ 是在第 k 票刚投了以后，候选人 P 所領先或落后的票数，显然 $s_x = p - q$，而且

$$s_i - s_{i-1} = \varepsilon_i = \pm 1, \ s_0 = 0, \ (i = 1, 2, \cdots, x). \quad (1.1)$$

反过来,滿足条件(1.1)的每一个整数的排列$\{s_1, s_2, \cdots, s_x\}$都表示一个可能的投票記录. 我們将要采用几何的語言,用一条其第 i 边之斜率为 ε_i 第 i 个頂点之縱坐标为 s_i 的折綫来表示这样一个排列. 这种折綫我們叫作路径.

定义. 設 $x > 0$ 和 y 都是整数. 从原点到点(x, y)的路径$\{s_1, s_2, \cdots, s_x\}$就是一条折綫,其頂点的横坐标为 $0, 1, 2, \cdots, x$,其頂点的縱坐标 $s_0, s_1, \cdots, s_x$ 滿足(1.1),而且 $s_x = y$.

如果 ε_i 中有 p 个正的 q 个負的,則

$$x = p + q, \ y = p - q. \quad (1.2)$$

如果能用一条路径把(x, y)与原点联結起来，那么 x, y 必是形如(1.2)的数. 这时，从 $x = p + q$ 个可能位置中选取 p 个正 ε_i 的位置的不同选法共有

$$N_{x,y} = \binom{p+q}{p} = \binom{p+q}{q} \quad (1.3)$$

个. 为了方便起見，当 x, y 不是形如(1.2)的数时，我們定义 $N_{x,y}=0$. 于是,联接原点和(x, y)的路径共有 $N_{x,y}$ 条. 伯川的选举問題的定理断言：当 $y > 0$ 时，滿足条件 $s_1 > 0, s_2 > 0, \cdots, s_{x-1} > 0, s_x = y$ 的路径共有 $(y/x)N_{x,y}$ 条. 此定理将在第 2 节中

---

1) 关于这个問題的历史和文献参看[21]，[22].

証明.

例. 图 1 描出了一条到 $N_1=(5,1)$的路径. 这样的路径共有 10 条,其中恰有两条满足 $s_i>0$. 这两条路径中的一条就是图 1 所描繪的$\{1,2,1,2,1\}$,而另一条就是$\{1,2,3,2,1\}$.



图 1. 正路径和第 2 节定理 2 的証明的示意图.

如果我們取消路径的終点 (x, y) 事先固定的这个約定,那么便可以从选举問題的定理推出很多有兴趣的結論.从原点到(n,y)(纵坐标 y 任意)的路径共有 2^n 条. 如第 3 节所解释的, 这 2^n 条路径可以解释为連續扔一个均匀的錢币 n 次所得到的 2^n 个可能的結果. 經典的描述是引进一个假想的賭徒甲, 他在每一次試験时或者赢一元,或者輸一元. 因此, 序列$\{s_1, s_2, \cdots, s_n\}$代表甲的接連的累积赢利,也就是正面超过反面的累积次数.

如果 $s_n=0$, 即是第 n 次試驗刚作完后其赢利为 0, 即是一次和局. 和局的出現是非常少見的,以至于它們不会影响图形,但是,重复地提到它是很麻烦的. 因此,我們約定当 $s_n>0$ 时,或者当 $s_n=0$ 且 $s_{n-1}>0$ 时(即是当和局出現时,誰在前一次試験領先,誰就被認为是領先),称甲領先. "甲在第 n 次試驗領先"与"第 n 边在 x 軸上方"的意义是一样的.

选票問題是考虑一些完全处于 x 軸上方的路径,即是,考虑这样一个賭博問題,其領先权是永不变的. 这个論題的进一步的研究是研究任何一条路径的領先权改变的頻繁程度. 与此联系,我們得到了一些与直观相差很远的結論. 一般都这样想: 在一长串扔錢币的序列中,甲,乙領先之次数各半. 然而,这是完全錯誤的. 在 20,000 次抛擲中,某甲在 20,000 次中一直領先的可能性比甲乙各領先 10,000 次的可能性約大 88 倍. 一般地說, 領先权改变得

如此不頻繁致使它是对直观想象的一个挑战. 不管抛擲的序列有多长,領先权改变的次数的最大可能的数値是 0; 領先权恰巧改变一次的可能比改变两次的可能大, 改变两次的可能比改变三次的可能大,等等. 簡而言之,如果一个現代的教育家或者心理学家去描写个别的扔錢币的賭博中长久的历史, 他一定会把絕大多数錢币視为不均匀的不完整的. 如果把很多錢币每个扔 n 次,它們之中有非常惊人大的一部分使領先权几乎在全部时間中都落在某一个賭徒手中;而有很少一部分使領先权換边,并且它的起伏变化正象人們对标准錢币一般所期望的那样.

这是由第一反正弦律(参看第 5 节和第 7 节的解释)所导出来的一个結論. 安德生曾經指出: 这个定律的应用范围是很广泛的, 这里所描述的关于扔錢币的情形是累积效应的随机起伏的典型問題. 物理学中,經济学中,教育学中很多随机过程都是这类性質的, 而且我們的发現对那些执恋于分辨通常趋势和偏离平均的人来說是一个忠告.

可以从各种不同的角度来考虑同样的情形. 如果用一致的速度扔錢币, 一般的想象总是这样: 适当的承認随机起伏,作两天試驗所出現的和局的次数应为作一天試驗所出現的和局的次数的二倍. 換句話說,直观上,我們期望和局的次数与試驗的持續时間成比例地增加. 事实上并不如此,和局的次数大致上与試驗的持續时間的平方根成比例地增加. 在 10,000 次抛擲中, 和局次数的中位数为 67, 但是,在 1,000,000 次抛擲中, 和局次数的中位数增为 674; 典型"波长"約从 150 增为 1500. 平均"波长"随时間之增加而增加(第 6 节和第 8 节). 作为这些結論的基础的公式在一般的随机徘徊和扩散理論中的初过时間和再現时間中起着很重要的作用.

第 2 节定理 3 和其余的內容沒有联系,而且在其它的地方也用不着. 它是两个候选人都获得同样多的票数(譬如說n)的选举問題的变化問題. 这时候选人甲領先 $2k$ 次,而乙在其余的 $2n-2k$ 次領先. 这里,我們又会产生錯覚,以为每个候选人領先的次数大約各为一半, 卽是,我們期望 $2k$ 与 n 很靠近. 其实不然, 如果选举的結果是和局 $n:n$, 則領先权的 $n+1$ 种可能分配(为 $2n:0$, $2n-2:2$, $2n-4:4$, $\cdots$, $2:2n-2$, $0:2n$) 都具有同样的概率 $(n+1)^{-1}$. 这个結果与前面所描述的那种对最后結果不事前給定的情形形

成一个强烈的对比；在那里最极端的分配 $2n:0$ 和 $0:2n$ 都是最大可能的。

郝捷士(J. L. Hodges)曾經指出[23]： 这些定理在“行序”检驗中有統計应用。我們用下面的例子来解释这点。

例. 假定对 n 个精心培植的东西的每一个都测得一个量，同时也对另外 n 个被控制的东西的每一个都测得一个量(例如植物的高)，从而得到数据 $a_1, \cdots, a_n$，和 $b_1, \cdots, b_n$。 为了确定起見，假定每一組都是按下降的次序排列的： $a_1 > a_2 > \cdots$ 和 $b_1 > b_2 > \cdots$。 让我們把这两个序列結合在一起，并把这 $2n$ 个字母 $a_1, \cdots, b_n$ 按下降的次序写出。 n 个 a 和 n 个 b 排列的結果可以解释为一个选举問題，其中每一个候选人都得到 n 张选票。一种非常成功的处理是所有的 a 都在 b 的前面；一种完全无效的处理是它們构成一个随机的次序。在我們的排列中，如果在有 k 个不同的 a 在其相应的 b 的前面，即是如果不等式 $a_i > b_i$ 对 k 个指标成立，則說 a 恰巧領先 $2k$ 次。如果假定随机性，則这件事情发生的概率为 $1/(n+1)$，因此，a 領先 $2k$ 或者多于 $2k$ 次的概率为 $(n-k+1)/(n+1)$。 关于这个討論的典范(沒有概率論知識只是定性的)是由格尔頓(Galton)所給出的，他在 1876 年用之于却尔斯·达文(Charles Darwin)給他的資料上。在他的例子中 $2n$ 是 30 且 a 領先26次。格尔頓断言，这种处理仍然是有效的，但是在只有随机性的假定下，甚至一种无效的处理，在 16 次类似的試驗中有三次可能会产生領先次数超过 26。这就說明： 定性的分析对我們的不可靠的直观来說，仍然是一种有价值的补充。(至于一些有关的检驗，是基于連貫理論上的。参看第二章 5a。)

2. 排列問題

令 $A=(a, \alpha)$和 $B=(b, \beta)$为正象限內的整点： $b > a \geqslant 0$，$\alpha > 0, \beta > 0$。点 A 关于 x 軸的反射点就是点 $A'=(a, -\alpha)$(参看图 2)。从 A 到 B 的路径的定义仍如 §1，不过这里 A 点相应于那里

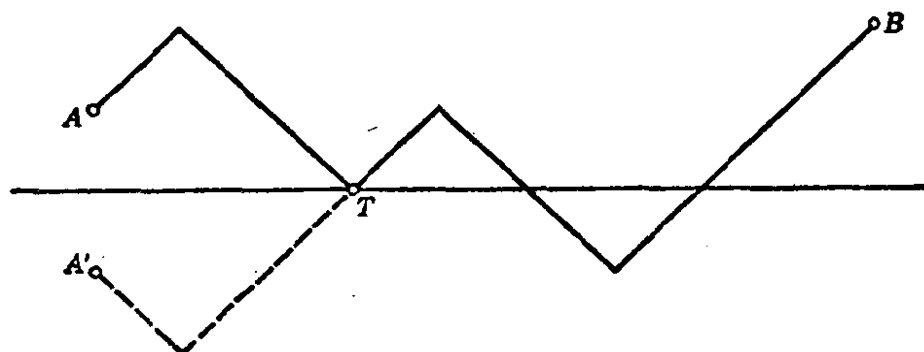


图 2. 反射原理的示意图。

的原点.

引理[1](反射原理). 从A到B的路径中触到或者穿过x軸的路径的个数等于从A'到B的路径的个数.

証. 考虑一个从A到B且在x軸上有一个或一个以上的頂点的路径$\{s_a=\alpha, s_{a+1}, \cdots, s_b=\beta\}$. 令$t$为第一个这样的頂点的横坐标(参看图 2);也就是說:选取t,使得$s_a>0, \cdots, s_{t-1}>0$, $s_t=0$. 因此,$\{-s_a, -s_{a+1}, \cdots, -s_{t-1}, s_t=0, s_{t+1}, s_{t+2}, \cdots, s_b\}$是引导$A'$到$B$的路径,而且$T=(t, 0)$是它在$x$軸上的第一个頂点. 折綫$AT$和$A'T$互为反射,从而全部从$A$到$B$的有一頂点在$x$軸上的路径与全部从$A'$到$B$的路径之間存在一个一一对应. 这就証明了我們的引理.

定理 1(选举定理). 令$x>0$, $y>0$; 从原点到(x, y)使$s_1>0, s_2>0, \cdots, s_x>0$的路径$\{s_1, s_2, \cdots, s_x=y\}$的总数为$(y/x)N_{x,y}$.

証. 因为$s_1=\pm 1$,所以对于每一个合乎定理要求的路径都有$s_1=1$. 这就推出:合乎定理要求的路径的数目与从(1,1)到(x, y)的而且旣不触到x軸也不通过x軸的路径数相等. 由上面的引理得知,这样的路径的数目为

$$N_{x-1,y-1}-N_{x-1,y+1}=\binom{p+q-1}{p-1}-\binom{p+q-1}{q-1}=$$

$$=\frac{p-q}{p+q}\binom{p+q}{p}=\frac{y}{x}N_{x,y}. \tag{2.1}$$

对偶原理. 几乎每个关于路径的定理都可以改变陈述而得到形式上不同的定理. 考虑$\{s_1, \cdots, s_n\}$和由它顚倒ε_i的次序而得到的路径$\{s_1^*, s_2^*, \cdots, s_x^*\}$,其中

$$s_1^*=\varepsilon_x, s_2^*=\varepsilon_x+\varepsilon_{x-1}, s_3^*=\varepsilon_x+\varepsilon_{x-1}+\varepsilon_{x-2}, \cdots$$

---

1) 在概率文献中認为此方法是由安德(D. Andre)所給. 本书把它化为随机徘徊的一个引理. 古典的随机徘徊(第十四章)的差分方程与微分方程非常类似,而反射原理(甚至它的更強的形式)在微分方程中熟知为凱尔文(Kelvin)勛爵的映象方法.

或者

$$s_0^* = 0,\ s_1^* = s_x - s_{x-1},\ \cdots,\ s_i^* = s_x - s_{x-i},\ \cdots,\ s_x^* = s_x. \tag{2.2}$$

路径(1.1)和(2.2)是重合的，而且它們之中的任一个可由另一个旋轉 180° 而得到；它們所联結的頂点是相同的。对于路径的每一个定理，都可以对应着一个把它应用到倒路径(2.2)去而得的对偶定理。

例如，选举定理給出了从原点到(x, y)而且 $s_i^*>0 (i = 1, 2, \cdots, x)$的倒路径$\{s_1^*, \cdots, s_x^*\}$的数目。这个数目与满足 $s_x > s_{x-i}$ $(i = 1, 2, \cdots, x-1)$的数目相等，因而，我們得到了选举定理的另一形式：

定理 1*. 从$(0,0)$到(x, y)的满足 $s_1<s_x, s_2<s_x, \cdots, s_{x-1}<s_x$（其中 $s_x = y>0$）的路径数等于$(y/x)N_{x,y}$.

如果用几何的語言来說，定理 1 中所考虑的路径是左端点为最低頂点的路径，而定理 1* 所考虑的路径是右端点为最高的頂点的路径（参看图 3）。定理 1* 与随机徘徊的初过时間有联系。

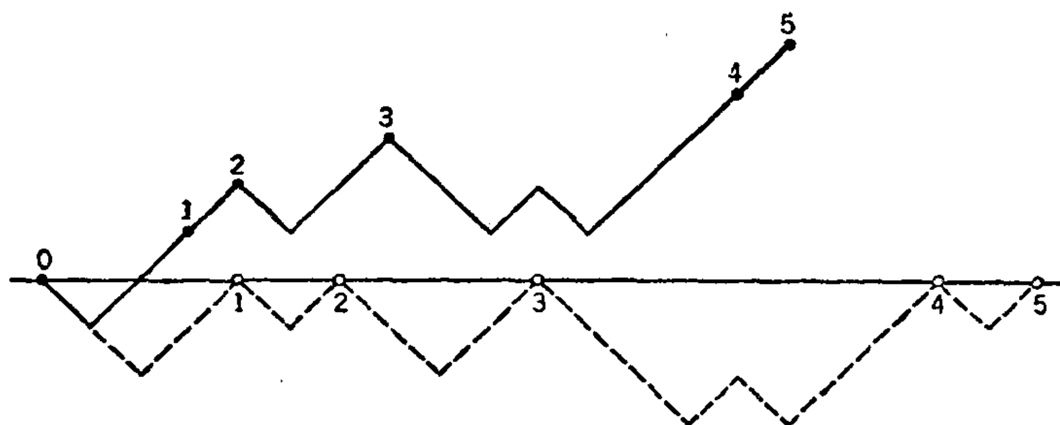


图 3. 初过时間和返回原点的示意图.

現在我們再回过来討論联結原点和 x 軸上的点 $N = (2n, 0)$ 的路径（在 x 軸上的端点数不可能是奇数）。令

$$L_{2n} = \frac{1}{n+1}\binom{2n}{n}. \tag{2.3}$$

定理 2. 在联結原点和 x 軸上的点 $2n$ 的$\binom{2n}{n}$条路径中

(a) 恰有 L_{2n-2} 条路径满足

$$s_1 > 0,\ s_2 > 0,\ \cdots,\ s_{2n-1} > 0\ (s_{2n} = 0), \tag{2.4}$$

(b) 恰有 L_{2n} 条路径满足

$$s_1 \geqslant 0,\ s_2 \geqslant 0,\ \cdots,\ s_{2n-1} \geqslant 0\ \ (s_{2n} = 0). \tag{2.5}$$

(这就是說，由原点到 $2n$ 的全部內頂点都在 x 軸以上的路徑数与由原点到 $2n-2$ 的沒有頂点在 x 軸以下的路徑数相等.)

証(参看图 1). 每一条满足条件(2.4)的路徑都通过点 $N_1=(2n-1,1)$,因此,由定理 1 得知,到点N_1且满足$s_1>0,\cdots,s_{2n-2}>0$的路徑数为

$$\frac{1}{2n-1}\binom{2n-1}{n-1} = \frac{1}{n}\binom{2n-2}{n-1} = L_{2n-2}. \tag{2.6}$$

这就証明了(a). 再令一条路徑满足条件(2.4). 截去其最初和最后一边,我們就得到了一条联結 $O_1=(1,1)$和 $N_1=(2n-1,1)$的而且其全部頂点或則在直綫 $y=1$ 的上方或則落在 $y=1$ 上的路徑. 把原点移至 O_1 后,我們就得到了一条联結新的原点和 N_1(它的两个新坐标是 $2n-2$ 和 0)的路徑. 它沒有一个頂点在新的 x 軸下面. 因此，我們在这样的路徑与满足条件(2.4)的路徑之間建立了一个一一对应,这就証明了我們的定理.

如第 1 节所說明的,下面的定理与其余的无联系,而且以后也用不着.

定理 3[24]. 令 $L_{2k,2n}$ 为从原点到 x 軸上的点 $2n$ 的恰有 $2k$ 条边在 x 軸上面而有 $2n-2k$ 条边在 x 軸下面的路徑的数目($k=0,1,\cdots,n$). 于是 $L_{2k,2n}=L_{2n}$ (不依賴 k).

証. 当 $n=1$ 时,結論显然成立. 我們用归納法証. 假定 $L_{2k,2\nu}=L_{2\nu}$ 对 $\nu=1,2,\cdots,n-1$; $0\leqslant k\leqslant \nu$ 成立. 我們想要数出$\{s_1,s_2,\cdots,s_{2n}=0\}$这些路徑中恰有 $2k$ 个边在 x 軸上方的路徑个数.首先我們假定$1\leqslant k\leqslant n-1$. 这样的路徑經过 x 軸，我們令此路徑在 x 軸上的第一个頂点的横坐标为 $2r$. 然后考虑二族路徑.

第一族中的每一条路徑在 0 到 $2r$ 之間是正的，而且它在 $2r$ 到 $2n$ 之間那一段路徑恰有 $2k-2r$ 条边在 x 軸之上方. 其中 $k\geqslant r$. 由定理 2(a)得知，路徑$\{s_1,s_2,\cdots,s_{2r-1},s_{2r}=0\}$中满足 $s_1>0,\cdots,s_{2r-1}>0$ 的共有 L_{2r-2} 条，而且,由归納法假設得知,联結$(2r,0)$和$(2n,0)$的同时恰有 $2k-2r$ 条边在 x 軸上方的路徑数为 $L_{2k-2r,2n-2r}=L_{2n-2r}$. 因此,这一族共有 $L_{2r-2}L_{2n-2r}$ 条路徑.

第二族中的每一条路徑在 0 与 $2r$ 之間都是負的;它在 $2r$ 与 $2n$ 之間那

一段恰有 $2k$ 条边在 x 軸的上方. 由上面的推理得知,这样的路径个数仍然是 $L_{2r-2}\ L_{2n-2r}$,不过这时 $n-r\geq k$.

由此推出,对于 $k=1,\cdots,n-1$, 我們有

$$L_{2k,2n}=\sum_{r=1}^{k}L_{2r-2}L_{2n-2r}+\sum_{r=1}^{n-k}L_{2r-2}L_{2n-2r}. \tag{2.7}$$

变换求和号中的附标: $\rho=n-r+1$, 上面的第二个和的项变为 $L_{2r-2}L_{2n-2r}=L_{2\rho-2}L_{2n-2\rho}$ 此时 ρ 由 $k+1$ 跑到 n. 因此

$$L_{2k,2n}=\sum_{\rho=1}^{n}L_{2\rho-2}L_{2n-2\rho}, \tag{2.8}$$

它不依賴 k.

全部 $2n$ 个边都在 x 軸上方的路径正是定理 2(b) 中所考虑的那种类型的路径,因此 $L_{2n,2n}=L_{2n}$. 由于对称性, 我們也有 $L_{0,2n}=L_{2n}$. 又因由原点到 $(2n,0)$ 的路径的总数为 $(n+1)L_{2n}$,故有 $L_{2k,2n}=L_{2n}$ 对 $k=0,1,\cdots,n$.

作为一个系,我們有

$$L_{2n}=\sum_{\rho=1}^{n}L_{2\rho-2}L_{2n-2\rho}. \tag{2.9}$$

此式的直接的分析証明請参看 8(a).

3. 随机徘徊和扔錢币

在一个 N 次扔錢币的序列中, 如果第 k 次出現正面, 則令 $\varepsilon_k=1$, 否則令 $\varepsilon_k=-1$. 于是 $s_k=\varepsilon_1+\varepsilon_2+\cdots+\varepsilon_k$ 是 n 次抛擲后正面比反面多出現的次数. 如果用賭博的語言来說, s_k 是"某甲所获之純利". 接連 N 次扔錢币中每一个可能的結果都可以用一条从原点开始的具有 N 边的路径来表示,反过来,每一条这样的路径都可以代表 N 次扔錢币的一个結果.

这种考虑使得我們可以把从原点出发的 2^N 条路径 $\{s_1,\cdots,s_N\}$ 作为我們的样本空間,而且每一条路径賦以概率 2^{-N}.

"最初两次試驗出現正面"这一事件必須解释为全部这样的序列: 其中 $s_1=1$, $s_2=2$. 这样的序列共有 2^{N-2} 个, 因此这一事件的概率为 2^{-2}. 更一般地,如果 $k<N$, 則恰有 2^{N-k} 条不同的路径 $\{s_1,s_2,\cdots,s_N\}$ 其最初 k 个頂点落在給定的路径 $\{s_1,\cdots,s_k\}$ 上. 这就推出: 由最初 $k(\leqslant N)$ 次試驗所决定的事件的概率不依賴于

N. 因此，在实际中，当N充分大时，N幷不起什么作用. 从概念上来看或者从形式上来看，最好是把一个抛擲的有限序列都考慮为一个非明显的无穷序列的前面一段，但是，这会要求我們考慮不可数的样本空間. 因此，我們将只考慮长为N的序列，其中N比公式中所出現的試驗的次数要大；此外，我們将乐意地忘却N.

为了加強概率背景及其与相关問題的联系，我們希望用不同术語来补充几何語言. 我們設想每隔一段固定的时間扔一次錢币，因此，第n次試驗在时刻n进行. 某甲可以用一个指示器(我們称它为"質点")标記每一个时刻的累积的贏利. 因此，这个質点从原点开始在垂直綫上运动. 它在时刻1, 2, ···运动，当錢币出現正面时它向上移一步，当錢币出現反面时，它向下移一步. 我們說：这个質点作对称的随机徘徊. (物理学家把它作为一維扩散的最簡单的模型；参看第十四章.)

在时刻n，質点的位置是垂直綫上的点s_n. 路径$\{s_1, \cdots, s_N\}$表示随机徘徊的空間——时間图解，x軸相当于时間軸.

由这个直观背景的指引，我們引进下述术語.

术語. 我們說在时刻n質点回到原点，如果$s_n = 0$. 說在时刻n質点第一次回到原点，如果

$$s_1 \neq 0, \ s_2 \neq 0, \ \cdots, \ s_{n-1} \neq 0, \ s_n = 0. \tag{3.1}$$

說在时刻n質点初过$r > 0$，如果

$$s_1 < r, \ s_2 < r, \ \cdots, \ s_{n-1} < r, \ s_n = r. \tag{3.2}$$

至于第二次，第三次，···回到原点，初过$r < 0$都可以仿照上面的办法定义. 注意，只能在偶数时刻分子才能回到原点，而且我們将常常把公式限制在偶数时刻. 用賭博的語言来說，返回原点表示正面出現的累积次数与反面出現的累积次数相等. (图3列出二条路径，其中初过和返回原点都分别地标出；第二条路径具有每一边都是負的特性.)

4. 組合定理的修正

在以后各节里，我們将要引用下列符号：

$$u_{2n} = \binom{2n}{n} 2^{-2n}, \quad n = 0, 1, 2, \cdots \tag{4.1}$$

和

$$f_0 = 0, \quad f_{2n} = \frac{1}{2n} u_{2n-2}, \quad n = 1, 2, \cdots. \tag{4.2}$$

容易驗証

$$f_{2n} = u_{2n-2} - u_{2n}, \quad n = 1, 2, \cdots. \tag{4.3}$$

定理 1. 对于每一个 $n \geq 1$, 都有

$$u_{2n} = \mathbf{P}\{s_{2n} = 0\}, \tag{4.4}$$

$$u_{2n} = \mathbf{P}\{s_1 \neq 0, s_2 \neq 0, \cdots, s_{2n} \neq 0\}, \tag{4.5}$$

$$u_{2n} = \mathbf{P}\{s_1 \geq 0, s_2 \geq 0, \cdots, s_{2n} \geq 0\}, \tag{4.6}$$

用話来説，下述三个事件的概率都是 u_{2n}:(a) 在时刻 $2n$ 質点返回原点;(b)一直到时刻 $2n$ 为止都沒有返回原点;(c)在 0 到 $2n$ 之間路径是非負的.

此外

$$f_{2n} = \mathbf{P}\{s_1 \neq 0, s_2 \neq 0, \cdots, s_{2n-1} \neq 0, s_{2n} = 0\}, \tag{4.7}$$

$$f_{2n} = \mathbf{P}\{s_1 \geq 0, s_2 \geq 0, \cdots, s_{2n-2} \geq 0, s_{2n-1} < 0\}, \tag{4.8}$$

这就是說,下面两个事件的概率都是 f_{2n}:(a) 在时刻 n 第一次返回原点;(b)在时刻 $2n-1$ 初过 -1.

証. 如第 3 节所討論的，只需考虑固定长度 $2n$ 的路径所构成的样本空間. 由(1.3)得知: 联結原点与 $(2n, 0)$ 的路径共有 $\binom{2n}{n}$ 条,这就証明了(4.4).

由第 2 节定理 2(a),联結原点和$(2n, 0)$的且满足 $s_1>0, \cdots, s_{2n-1}>0$ 路径的共有 L_{2n-2} 条. 因此,滿足条件(4.7)的路径的总数为上述总数的两倍,从而其概率为 $2L_{2n-2} \cdot 2^{-2n} = f_{2n}$. 用同样的办法,由第 2 节定理 2(b) 可得(4.8).

一直到时刻 $2n$ 为止,尙未返回原点的概率等于 1 減去在小于等于 $2n$ 时刻上第一次返回原点的概率. 应用(4.7),这个差为

$$\begin{aligned} & 1 - f_2 - f_4 - \cdots - f_{2n} = \\ & = 1 - (1 - u_2) - (u_2 - u_4) - \cdots - (u_{2n-2} - u_{2n}) = \\ & = u_{2n}, \end{aligned} \tag{4.9}$$

这就証明了(4.5). 类似地，(4.6)的右边等于1减去在时刻 $2n$ 以前初过 -1 的概率，应用(4.8)，这个差仍为(4.9). 这就完成了我們的証明.

系. 对 $n \geq 1$ 有

$$u_{2n} = \sum_{r=1}^{n} f_{2r} u_{2n-2r}. \tag{4.10}$$

証. 如果在时刻 $2n$ 返回原点，则第一次返回原点一定在某个时刻 $2r \leq 2n$. 我們刚才看到，联結原点与 $(2n, 0)$ 的路径而且在 $2r \leq 2n$ 时第一次返回原点的条数为 $2^{2r}f_{2r} \cdot 2^{2n-2r}u_{2n-2r}$. 对 r 求和，我們得到方程(4.10). （直接的分析証明参看8(a). 在第十三章第3节中，我們将要看到：(4.10)是再現事件的基本方程的特别情形.）

第2节中的定理1* 算出了在时刻 x 初过 y 的路径数. $x+y$ 必須是偶数，为了方便起見，令 $x = 2n - y$. 于是定理1* 的內容可以表述如下：

定理2. 在时刻 $2n-y$ 初过 $y>0$ 的概率为

$$f_{2n}^{(y)} = \frac{y}{2n-y}\binom{2n-y}{n} 2^{-2n+y}, \quad n \geq y > 0. \tag{4.11}$$

用对偶原理把上述的重要公式作为选举定理的直接推論而推出，这种簡单的方法是值得注意的. 公式(4.11)的直接的分析推导是困难的，它需要一些特别的技巧.

原則上，概率 $f_{2n}^{(y)}$ 可以对 y 用归納法来算出. 一条长度为 $2n-y-1$ 的在終点初过 $y+1$ 的路径可以分割为二段(参看图3，$y=4$). 第一段是由原点到初过 y 那一点所构成的路径；初过 y 发生在某一个时刻 $2\nu - y < 2n - y - 1$. 継第一段之后是第二段，第二段是一条长为 $2n-2\nu-1$ 的、其終点是唯一的一个在左端点之上的頂点的路径. 換句話說，如果其左端点是原点，則第二段是一条在終点初过1的路径. 由定义得知：第一种类型的可能路径共有 $2^{2\nu-y} f_{2\nu}^{(y)}$ 条，第二种类型的可能路径共有 $2^{2n-2\nu-1} f_{2n-2\nu}^{(1)}$ 条，而且任何两条都可以联結成一条在时刻 $2n-y-1$ 初过 $y+1$ 的路径. 因此

$$f_{2n}^{(y+1)} = \sum_{\nu=y}^{n-1} f_{2\nu}^{(y)} f_{2n-2\nu}^{(1)}, \quad n \geq y+1. \tag{4.12}$$

公式(4.8)說明：在时刻 $2n-1$ 初过 -1 (因此初过 $+1$ 也一样)的概率为 f_{2n}，这就是說

$$f_{2n}^{(1)}=f_{2n},\ n\geqslant 1. \tag{4.13}$$

公式(4.12)和(4.13)遞推地决定了全部 $f_{2n}^{(y)}$，但是不容易驗証(4.11)满足(4.12)，因此，(4.11)这个明显公式如何由(4.12)推出是不显然的.

由公式(4.12)和(4.13)可以导出一个新奇的結論. 由(4.13)我們看出 $f_{2n}^{(1)}$ 是在时刻 $2n$ 初返原点的概率. 許我們忘掉上面的定理，我們定义 $f_{2n}^{(y)}$ 为在时刻 $2n$ 第 y 次返回原点的概率. 不加改变地应用上述証明中所进行的推导：把一条由原点到第$(y+1)$次返回原点的路径分为二段，前一段为原点到第 y 次返回原点，后一段为其余的，我們将再一次地看出(4.12)成立. 因为这个关系唯一地决定了全部的 $f_{2n}^{(y)}$，所以我們有

定理 3. 在时刻 $2n$ 第 y 次返回原点的概率由(4.11)所給出.

另外的几何証明. 考虑一条由原点到时刻 $2n-y$ 初过 y 的路径.（图 3 显示了 $y=5$，$2n-y=15$ 的情形.）造一条新的路径如下：把原来的路径再插入 y 个新的边，每一条新边的斜率都是 -1，这 y 条边的左端点分别为原点和 $y-1$ 个这样的頂点，在它們上面分別初过 $1, 2, \cdots, y-1$. 这条新的路径我們以$\{\sigma_1, \sigma_2, \cdots, \sigma_{2n}\}$表之，它的长度为 $2n$. 显然，$\sigma_1\leqslant 0, \cdots, \sigma_{2n-1}\leqslant 0$，$\sigma_{2n}=0$，而且恰恰有 $y-1$ 个內頂点落在 x 軸上. 反之，每一条具有这样性質的路径，可以从在 $2n-y$ 时刻初过 y 的路径用上述方法求得. 令 $f_{2n}^{(y)}$ 为定理 2 中所定义的概率(即在 $2n-y$ 时刻上初过 y 的概率)，我們看到具有下述性質的路径：$\{\sigma_1, \cdots, \sigma_{2n}\}$，$\sigma_i\leqslant 0$，$\sigma_{2n}=0$ 且恰有 $y-1$ 个內頂点在 x 軸上，共有 $2^{2n-y}f_{2n}^{(y)}$ 条. 这样一条路径是由 y 段所构成，其端点在 x 軸上，因此，当我們变換这种路径的一段或几段中所有 σ_i 的符号的时候，我們可以得到 2^y 条不同的路径. 用这种办法，我們得到了全部长为 $2n$ 的满足 $s_{2n}=0$ 且恰有 $y-1$ 个內頂点落在 x 軸上的路径，因此，它們的总数为 $2^{2n}f_{2n}^{(y)}$，这就証明了我們的論断.

5. 长的領先概率：第一反正弦律

我們說質点在时刻 $k-1$ 到 k 处于其运动路径的正边上，如果其路径的第 k 边在 x 軸的上方，也就是說，s_{k-1}，s_k 这两个頂点中至少有一个是正的(在这种情形下，其它一个一定是正的或者0). 用賭博的語言来說，这就意味着第 $k-1$ 次和第 k 次試驗某甲的累积贏利都是非負的.

在第 1 节中所談到的关于路径的反直观的性質可以由下述定

理推出.

定理 1[1). 令 $p_{2k,2n}$ 为从时刻 0 到 $2n$ 时质点有 $2k$ 个时間单位处于正边上而有 $2n-2k$ 个时間单位处于負边上的概率. 于是

$$p_{2k,2n}=u_{2k}u_{2n-2k}. \tag{5.1}$$

(注意: 在正边上的时間必須为偶数个时間单位.)

証. 在时間区間$(0, 2n)$中質点一直处于正边的概率由公式(4.6)給出,因此 $p_{2n,2n}=u_{2n}$, 这就証明了当 $k=n$ 时(5.1)成立. 由于对称性,我們还有 $p_{0,2n}=u_{2n}$, 因此, 只需对 $1\leqslant k\leqslant n-1$ 来証明(5.1)即可. 为此,我們再重复导出(2.7)的論証. 如果質点恰有 $2k>0$ 个时間单位处于正边, $2n-2k>0$ 个时間单位处于負边,則質点一定要通过零点. 令質点在时刻 $2r$ 第一次返回零点,则这条路径必須属于下列两族路径中的一族.

在第一族中,到时刻 $2r$ 为止,質点一直处于正边上,而在时間区間 $2r$ 到 $2n$ 中質点恰有 $2k-2r>0$ 个时間单位处于正边. 长为 $2r$ 的在时刻 $2r$ 第一次返回原点的路径数共有 $2^{2r}f_{2r}$ 条,其中的一半保持在正边. 此外,由定义得知: 由$(2r,0)$出发的长为 $2n-2r$ 的恰有 $2k-2r$ 个边在 x 軸上方的路径数为 $2^{2n-2r}p_{2k-2r,2n-2r}$. 因此,第一族中长为 $2n$ 的路径总数为

$$\frac{1}{2}\cdot 2^{2r}f_{2r}\cdot 2^{2n-2r}p_{2k-2r,2n-2r}=2^{2n-1}f_{2r}p_{2k-2r,2n-2r}.$$

在第二族中,在时刻 0 到 $2r$ 質点一直处于負边, 而在时刻 $2r$ 到 $2n$ 中質点恰有 $2k$ 个时間单位处于正边, 其中 $2k\leqslant 2n-2r$. 上面討論可知,这一族中的路径总数为 $2^{2n-1}f_{2r}p_{2k,2n-2r}$.

由此推出,对于 $1\leqslant k\leqslant n-1$ 有

$$p_{2k,2n}=\frac{1}{2}\sum_{r=1}^{k} f_{2r}p_{2k-2r,2n-2r}+\frac{1}{2}\sum_{r=1}^{n-k}f_{2r}p_{2k,2n-2r}. \tag{5.2}$$

用归納法, 假設 $p_{2k,2\nu}=u_{2k}u_{2\nu-2k}$ 对 $\nu=1, 2, \cdots, n-1$ 成立(这个关系对 $\nu=1$ 成立是不足道的). 于是公式(5.2)化为

1) 第一次用分析方法証明这个定理的是鍾开萊和費勒[24],而問題是在斯伯·安德生的工作中所提出的,参看 69 頁脚注 2).

$$p_{2k,2n}=\frac{1}{2}u_{2n-2k}\sum_{r=1}^{k}f_{2r}u_{2k-2r}+\frac{1}{2}u_{2k}\sum_{r=1}^{n-k}f_{2r}u_{2n-2k-2r}. \quad (5.3)$$

由方程(4.10)得知,第一个和等于 u_{2k},而第二个和等于 u_{2n-2k},这就証明了(5.1).

直观上,我們覚得处于正边的总时間与总时間之比 k/n 可能是近似地为$\frac{1}{2}$. 然而,事实恰恰相反: k/n 靠近$\frac{1}{2}$的概率很小,而两个极端情形 $k/n=0$ 和 $k/n=1$ 的概率却很大. 这个結論可以用商检験于(5.1)上来験証.

表 1 說明这个反直观現象. 用賭博的語言来說, 这結論揭示了一个使人惊奇的事实: 在扔 $2n=20$ 次均匀的錢币中, 不幸运的賭徒沒有一次領先的概率为 0.3524. 在較多的情况下(具有概率 0.5379), 不幸运的賭徒的累积赢利只有一次或一次都沒有为正的. 为了对比起見,領先的次数之比为 10:10 的概率只有 0.0606.

表 1. 在扔 20 次錢币中,領先的次数的分布

	$k=0$ $k=20$	$k=2$ $k=18$	$k=4$ $k=16$	$k=6$ $k=14$	$k=8$ $k=12$	$k=10$
$p_{k,20}=$	0.1762	0.0927	0.0736	0.0655	0.0617	0.0606
$P_{k,20}=$	0.3524	0.5379	0.6851	0.8160	0.9394	1

$p_{k,20}=u_k u_{20}$ 是路径中恰有 k 边在 x 軸上方的概率,即是"某甲在 20 次试験中恰有 k 次領先"的概率.

$P_{k,20}$ 是賭徒中之一至少有 k 次領先,而另一个至多只有 $20-k$ 次領先的概率.

公式(5.1)虽然是准确的,但是不够明朗, 因此最好用一个簡单的逼近来代替它. 应用第二章的史特令公式可証: 当 $n\to\infty$ 时 $u_{2n}(\pi n)^{\frac{1}{2}}\to 1$ [这就是第二章問題 (12.20) 的內容]. 由此推出

$$p_{2k,2n}\sim\frac{1}{\pi k^{\frac{1}{2}}(n-k)^{\frac{1}{2}}}, \quad (5.4)$$

(5.4)中两边之比当 $k\to\infty$, $n-k\to\infty$ 时迅速地趋于 1. 处于正边的总时間与总时間之比 k/n 落在$\frac{1}{2}$ 与 $\alpha\left(\frac{1}{2}<\alpha<1\right)$之間的

概率由

$$\sum_{\frac{1}{2}n<k<\alpha n} p_{2k,2n} \sim \frac{1}{\pi n} \sum_{\frac{1}{2}n<k<\alpha n} \left\{\frac{k}{n}\left(1-\frac{k}{n}\right)\right\}^{-\frac{1}{2}} \tag{5.5}$$

給出。其右边是就是逼近下列积分的黎曼和：

$$\pi^{-1}\int_{\frac{1}{2}}^{\alpha} \frac{dx}{\{x(1-x)\}^{\frac{1}{2}}} = 2\pi^{-1}\text{arc}\sin\alpha^{\frac{1}{2}} - \frac{1}{2}. \tag{5.6}$$

由于对称性，$k/n \leqslant \frac{1}{2}$ 的概率当 $n\to\infty$ 时趋于 $\frac{1}{2}$. 把这个概率加到(5.5)去，我們得到

定理2[1]. (第一个反正弦律)对于固定的 $\alpha(0<\alpha<1)$，处于正边的总时間与总时間之比 $k/n<\alpha$ 的概率当 $n\to\infty$ 时趋于

$$\pi^{-1}\int_{0}^{\alpha} \frac{dx}{\{x(1-x)\}^{\frac{1}{2}}} = 2\pi^{-1}\text{arc}\sin\alpha^{\frac{1}{2}}. \tag{5.7}$$

在实用上，公式(5.7)給出了一个很好的逼近，甚至对 n 的值

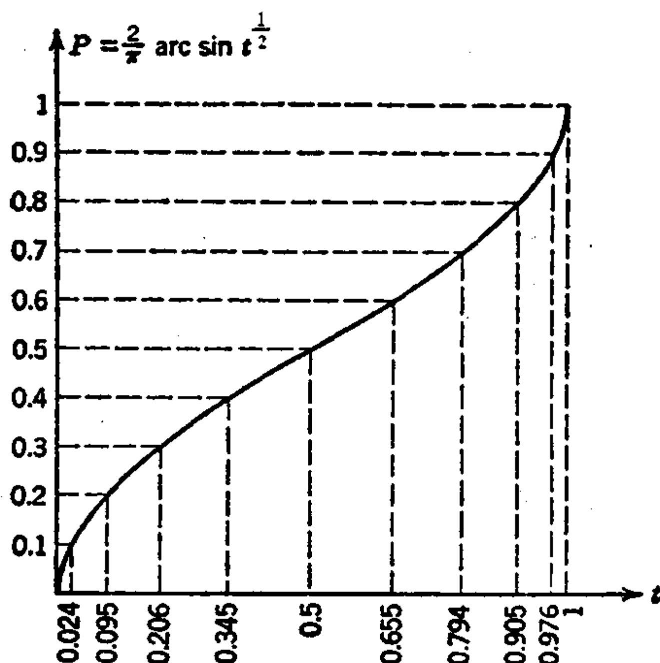


图 4. 反正弦律.

1) 保尔·列威 (Paul Lévy)在[25]中对給定的連續扩散过程找出了反正弦律，而且考虑了它和扔錢币的博奕的联系. 至于一般的相互独立随机变量序列的正的部分和的个数的反正弦律是由耳多斯(P. Erdös)和卡斯(M. Kac)所証明的，参看[26]. 至于反正弦律的組合性質及其对随机变量的一般情形的正确性是由斯伯·安德生所发現的.

在 20 左右，其逼近的程度都很好. (5.7) 中的被积函数的图形是一个U形曲綫,在其两个端点 0 和 1 处趋于无穷. 这就令人信服地表明了处于正边的总时間与总时間之比接近 0 或者 1 比接近"期望"或"正規"值$\frac{1}{2}$的可能性大. 图 4 将要表明:

处于原点同側的时間占总时間的 97% 的概率为 0.20. 十次中就有一次,質点以 99.4% 的时間处于原点同側. 表 2 給出另一种解释.

表 2. 反正弦律的解释

p	t_p	p	t_p
0.9	153.95 天	0.3	19.89 天
0.8	126.10 天	0.2	8.93 天
0.7	99.65 天	0.1	2.24 天
0.6	75.23 天	0.05	13.5 小时
0.5	53.45 天	0.02	2.16 小时
0.4	34.85 天	0.01	32.4 分

一个錢币,每秒鈡扔它一次,一直扔 365 天; 令 Z 为不幸运的賭徒領先的总时間. 于是 t_p 近似地是使事件 $Z<t_p$ 的概率为 p 的数.

这个表列出了一个不幸运的賭徒在全年中領先的时間少于 t_p 天的概率 p. 例如,統計学家一般喜欢取置信水平 $p=0.05$, 我們发現 20 次中就有一次使幸运的賭徒領先的时間大于 364 天零 10 小时. 很少有人会相信: 扔一个均匀的錢币会作成一个如此荒謬的进程,使得一百万次連續試驗中領先权永远不变,但是这正是一个好的錢币的規律性的結果.

下一节我們将要研究同一个現象的另外一类問題, 而第 7 节則是用經驗資料来解释理論.

6. 返回原点的次数

反正弦律的解释說明了这样一个事实: 質点返回原点之前需要进行很多次試驗. 几何地說,其路径越过 x 軸的次数很稀有.

直观上,我們覺得某甲和某乙扔一个錢币 $2n$ 次 (n 很大)时, 其和局(累

积赢利相等的时刻)的次数应該与 $2n$ 成正比。但是,事实并不如此。和局的次数仅仅依概率与 $(2n)^{1/2}$ 成比例;这就是說,把賭博的次数增大后,和局的頻率迅速下降,因而"波长"增加。为了分析这个結果,我們将要考虑返回原点的次数。必須記着:粗糙地說,質点由正边进入負边或者反之的次数为返回原点次数的一半。

定理 1. 令 $z_{2n}^{(r)}$ 为到时刻 $2n$ 为止質点共返回原点 r 次的概率。于是

$$z_{2n}^{(r)}=\frac{1}{2^{2n-r}}\binom{2n-r}{n},\qquad n\geq 1. \tag{6.1}$$

特别,$z_{2n}^{(0)}=z_{2n}^{(1)}=u_{2n}$,且

$$z_{2n}^{(0)}=z_{2n}^{(1)}>z_{2n}^{(2)}>z_{2n}^{(3)}>\cdots. \tag{6.2}$$

如果用文字来叙述,(6.2)說明:和賭博持續的时間 $2n$ 无关,不返回或者仅返回原点一次的概率比返回其它次数的概率大.

証. 由公式(4.4)和(4.5)得知,长为 2ν 的沒有一次返回原点的路径数与最后一步返回原点的路径数相等。現在我們来考虑一条长为 $2n$ 的在时刻 $2n-2\nu<2n$ 上第 r 次也是最后一次返回点的路径。从最后这次返回原点开始的长为 2ν 的路径数与从 $(2n-2\nu,\ 0)$ 出发到 $(2n,\ 0)$ 的路径数相等。換句話說,时刻 $2n$ 以前恰返回原点 r 次的概率与时刻 $2n$ 返回原点且在此以前曾至少返回原点 r 次的概率相等。由第 4 节定理 3,这就意味着

$$z_{2n}^{(r)}=f_{2n}^{(r)}+f_{2n}^{(r+1)}+f_{2n}^{(r+2)}+\cdots, \tag{6.3}$$

其中 $f_{2n}^{(y)}$ 由公式(4.11)給出。容易驗証

$$f_{2n}^{(y)}=\frac{1}{2^{2n-y}}\binom{2n-y}{n}-\frac{1}{2^{2n-y-1}}\binom{2n-y-1}{n}, \tag{6.4}$$

对 $y=r,\ r+1,\ \cdots$ 求和,我們就得到(6.1)。至于(6.2),那是一个不足道的推論,这就完成了我們的証明。

我們再用一个簡单的逼近来代替准确的公式(6.1)。为此,我們把(6.1)写作

$$z_{2n}^{(r)}=u_{2n}\frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{r-1}{n}\right)}{\left(1-\frac{1}{2n}\right)\left(1-\frac{2}{2n}\right)\cdots\left(1-\frac{r-1}{2n}\right)}. \tag{6.5}$$

正如反正弦律的証明中所指出的,当 $n\longrightarrow\infty$ 时我們有 $u_{2n}(\pi n)^{1/2}\longrightarrow 1$。由对数的泰勒展式得知 $\log\left(1-\frac{\nu}{n}\right)$ 可以用 $-\frac{\nu}{n}$ 来逼近,其誤差的阶为 $(\nu/n)^2$。由此我們有下述漸近式(其誤差的大小的阶为 r^3/n^2):

$$\log\{z_{2n}^{(r)}\pi^{1/2}n^{1/2}\}\approx-\frac{1}{2n}\sum_{\nu=1}^{r-1}\nu\approx-\frac{r^2}{4n}\tag{6.6}$$

或者

$$z_{2n}^{(r)}\approx\pi^{-1/2}n^{-1/2}e^{-r^2/4n}.\tag{6.7}$$

返回原点的次数小于 k 的概率 $z_{2n}^{(0)}+z_{2n}^{(1)}+\cdots+z_{2n}^{(k-1)}$ 可以由 $\pi^{-1/2}e^{-1/4x^2}$ 从 0 到 $\frac{k}{\sqrt{n}}$ 的黎曼和来逼近，相对(百分比)誤差的阶小于 k^3/n^2. 因此，我們有

定理 2. 对于每一个固定的 $\alpha>0$，直到时刻 $2n$ 为止，質点返回原点的次数少于 $\alpha(2n)^{1/2}$ 的概率当 $n\longrightarrow\infty$ 趋于[1)]

$$f(\alpha)=\left(\frac{2}{\pi}\right)^{1/2}\int_0^\alpha e^{-\frac{1}{2}s^2}\,ds.\tag{6.8}$$

特别，对充分大的 n 来說，返回次数少于 $0.6745(2n)^{1/2}$ 的概率漸近地为1/2.

在第七章第 1 节中，讀者将要看到正态分布函数 $\Phi(\alpha)=\frac{1}{2}\{1+f(\alpha)\}$ 的表，由这个表，可以算出 $f(\alpha)=2\left\{\Phi(\alpha)-\frac{1}{2}\right\}$(对 $\alpha>0$).

把一个錢币扔 10,000 次：返回原点的次数少于 68 的概率为 1/2，这其中有一半确实是改变領先权。换句話說，邻近的两次改变領先权之間的平均“波长”的約为 300 的概率是 1/2. 在扔 1,000,000 次錢币中，返回的中位数只增加 10 倍，因此平均波长的增到大約等于 3000. 試驗序列愈长，返回原点的次数愈少，因而波长愈长.

把一个錢币扔 10,000 次，不改变領先权的概率大約为 0.0085；在 1,000,000 次拋擲中，改变領先权的次数少于 10 的概率大約也为 0.0085.

7. 一个实驗的說明

图 5 代表把一个錢币扔 10,000 次所得到的实际的結果；它是第一章例(6.c)中所列的表的材料. 最上面那条綫包含了前 550次試驗的图形，而下面两条綫代表 10,000 次試驗的全部图形，只不过这时沿 x 軸方向的单位縮小了一些. 这两个图形中沿 y 軸方向的单位都是一样的.

---

1) 提醒熟悉中心极限定理的讀者：返回原点的次数不遵循正态分布. 在(6.8)中出現的是一个具有均值为 $(2/\pi)^{1/2}$ 及方差为 $1-(2/\pi)$ 的截尾的正态分布.

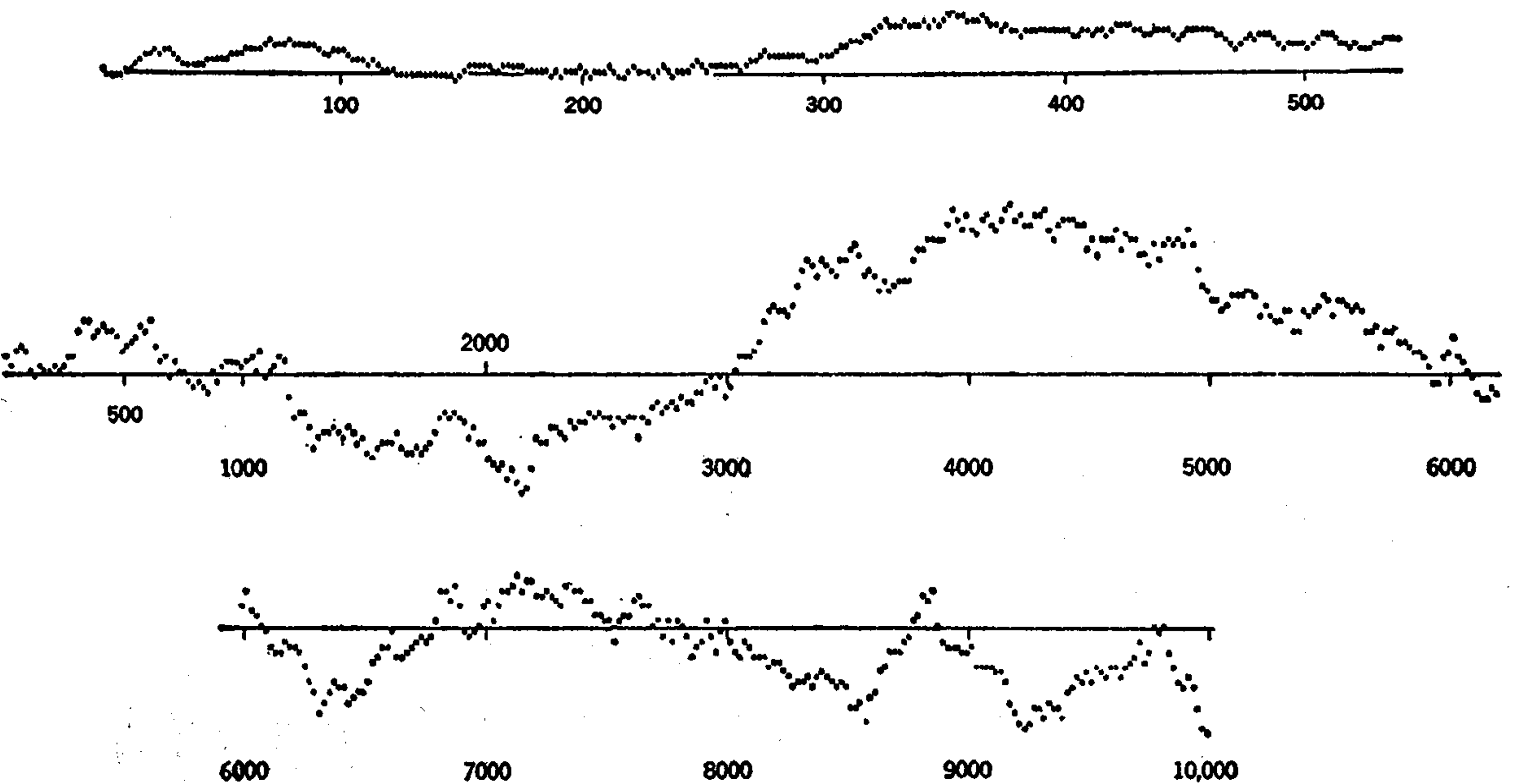


图 5. 把一个均匀的錢币扔 10,000 次所得到的記录(如第 7 节所描述).

看到这个图形以后，很多人会对邻近的两次越过 x 軸(即是改变領先权)之間的波长是如此的长而感到惊奇. 然而，这个图形表示的还是一个比較正常的情形，它是从三个可用的記录中选取其最正常的那一个. 讀者可以沿相反的方向看看这个图形，也就是說把終点当作原点. [分析上，反路径由 (2.2)給出.]理論上，图形中所描繪的点列与反向的点列是等价的，每一个都代表一个随机徘徊. 反向的随机徘徊具有下述特性. 从原点出发

"原点"停留在

負	边	正	边
首先	7804 步	其次	8 步
其次	2 步	其次	54 步
其次	30 步	其次	2 步
其次	48 步	其次	6 步
其次	2046 步		
总数为	9930 步	总数为	70 步
时間比例:	0.9930	时間比例:	0.007.

这似乎是荒謬的，但是，把一个均匀的錢币扔 10,000 次，領先权在一边的次数大于 9930 次而另一边的領先权少于 70 次的概率比 0.1 还要大一点. 换句話說，平均来說 10 次記录中就会有一次比刚才所描述的还要坏. 为了对比起見，領先权之差数比图 5 所描述的要好一些的記录的概率是很小的，大約为 0.072.

图 5 的記录中，包含了 142 次返回原点，其中有 78 次改变領先权. 上面所考虑的反向点列只包含 14 次返回原点，其中有 8 次改变領先权. 这个奇妙的例証所启示的內容，使得一个有修养的統計学家都觉得扔 10,000 次錢币只有 142 次正反相等是一个惊人小的数目，至于只有 14 次那就更出乎意外了 实际上正反相等次数大于 140 的概率大約为 0.157，而小于14的概率为 0.115. 因此，与直观矛盾的，只有 14 次正反相等就完全不觉得惊奇了；在考虑改变領先权的数目的当时，我們还考虑过反向的点列，它与图 5 中的原始点列处于同一地位.

8. 附　　录

(a) 一些等式的分析証明. 容易驗証

$$u_{2n} = (-1)^n \binom{-\frac{1}{2}}{n}, \quad f_{2n} = (-1)^{n+1} \binom{\frac{1}{2}}{n}. \tag{8.1}$$

基本等式(4.10)可以考虑作第二章(12.9)当 $a = \frac{1}{2}$, $b = -\frac{1}{2}$的特殊情形. 用同一个公式(求和)可得 $\sum_{r=0}^{n} u_{2r}u_{2n-2r} = 1$.

公式(2.8)可以用 f_{2k} 代替 L_{2k+2} 来改写,而且可推出第二章(12.9)中 $a = b = \frac{1}{2}$ 的特殊情形. 另外,公式(2.8)可以由公式(4.10)利用等式 $nr^{-1}(n-r)^{-1} = r^{-1} + (n-r)^{-1}$ 来推出.

(b) 最大值的位置: 第二反正弦律. 我們說路径 $\{s_1, s_2, \cdots, s_x\}$的第一个最大值在位置 k, 如果

$$s_k > 0, s_k > s_1, \cdots, s_k > s_{k-1}, s_k \geqslant s_{k+1}, \cdots, s_k \geqslant s_x. \tag{8.2}$$

特別,如果 $s_j \leqslant 0$ 对$1 \leqslant j \leqslant x$ 成立, 則第一个最大值在位置 0. 由公式(4.6)得知, 长为 $x = 2n$ 的路径的第一个最大值在位置 0 的概率为u_{2n}. 由此也推出, 对于一个长为 $x = 2n - 1$ 的路径, 其第一个最大值在 0 的概率为 u_{2n}.

事件“第一个最大值在最后一个位置”就是 $s_j < s_x$ 对 $j = 0, 1, \cdots, x-1$ 成立. 对于反向路径(2.2)来說, 这意味着 $s_1^* > 0$, $s_2^* > 0, \cdots, s_x^* > 0$, 这个概率由(4.5)給出, 它等于 $\frac{1}{2}u_{2n}$ (对 $x = 2n$ 或 $x = 2n+1$ 都对).

一条长为 $2n$ 的具有第一个最大值在位置 k 的路径由下述两段所組成: 第一段的第一个最大值在最后位置,即是在位置 k, 而第二段的第一个最大值在起点,即是在位置 0. 反过来, 任何具有上述性質的两段都可以組成一条路径, 其第一个最大值在位置 k. 因此,我們有

定理. 长为 $2n$ 的路径, 其第一个最大值的位置在 ν 的概率

等于

$$\begin{cases} \frac{1}{2}u_{2k}u_{2n-2k} & \text{当 } \nu = 2k \quad (k = 1, 2, \cdots, n), \\ & \text{或 } \nu = 2k + 1 \quad (k = 0, 1, \cdots, n - 1), \\ u_{2n} & \text{当 } \nu = 0. \end{cases} \tag{8.3}$$

值得注意的事实是：第一个最大值位于 $2k$ 或者 $2k+1$ 的概率等于(5.1)中的概率 $p_{2k,2n}$，$p_{2k,2n}$ 是质点在 $2n$ 个时間单位中有 $2k$ 个时間单位处于正边的概率。由此得知反正弦逼近式可以利用，从而我們断言：最大值位于两个端点中某一个的附近的傾向性很大。

領先的次数的分布 $\{p_{2k,2n}\}$ 与最大值的位置的分布实际上相同这件惊人的情形在扔錢币的賭博中是不奇怪的。斯伯·安德生对一大族随机变量証明了一个类似的定理，他的証明的組合基础与我們上面所应用的推理类似。

（c）**一个关于初过和返回原点的极限定理**[1]。第6节中所用到的估計可用来証明：对固定的 $y>0$，(4.11)中的概率 $f_{2n}^{(y)}$ 满足漸近关系

$$f_{2n}^{(y)} \sim \left(\frac{2}{\pi}\right)^{1/2} \frac{y}{(2n-y)^{3/2}} e^{-\frac{1}{2}y^2/(2n-y)}, \tag{8.4}$$

符号～表示两边之比当 $n \longrightarrow \infty$ 时趋于1。用来証明第5节和第6节中的极限定理的方法現在可以推导出下面的結論：第 y 次返回0（或者初过 y）的位置在时刻 ty^2 以前的概率当 $y \longrightarrow \infty$ 时趋于 $1 - f(t^{-1/2})$，其中 $f(\alpha)$ 由方程(6.8)所定义。

由此推出，第 y 次返回原点发生在时刻 $(2.21\ldots)y^2$ 以后的概率大約为 $1/2$，因此，邻近两次返回原点之間的平均时間依 y 粗略地綫性增加。这对于习惯于取 y 次"測同一量"的平均作为"眞正"值的近似的物理学家来說，是会感到惊奇的。在現在这种情形，进一步的分析就会告訴我們：在全部可能情形中，机会均等地 y 个测量中有一个的大小的阶与全部的和的大小的阶——y^2 一样。

1）这就是本书第一版第十二章第5节的定理3。以前我們告訴讀者 $1 - f(t^{-1/2})$ 是所謂1/2阶的正稳定分布。

*第四章[1)] 事件的組合

本章是討論关于由某些其他事件 $A_1, A_2, \cdots, A_N$ 所确定的一些事件. 例如在桥牌游戏中,“至少一家有同一花色”的事件 A 是由四个事件 $A_k(k=1,2,3,4)$ 所成之併, 其中 A_k 是事件“第 k 家有同一花色”. 由于諸事件 A_k 可能一个、两个或多个同时发生,也正因为有这些重迭关系,所以事件 A 的概率并不是四个概率 $\mathbf{P}\{A_k\}$ 之和. 对于給定的一組事件 $A_1, A_2, \cdots, A_N$, 我們将說明如何去計算其中有 0 个, 1 个, 2 个, ··· 同时发生的概率.

本章內容在弗雷謝(M. Fréchet)[27]的专著中可找到,如果讀者要作进一步的了解,可参閱这一专著.

1. 事 件 之 併

設 A_1 与 A_2 是两个事件, 則 $A = A_1 \cup A_2$ 表示 A_1 或 A_2 或两者同时发生的事件. 由第一章(7.4)式我們有

$$\mathbf{P}\{A\} = \mathbf{P}\{A_1\} + P\{A_2\} - P\{A_1A_2\}. \qquad (1.1)$$

現在, 我們希望将此公式推广到 N 个事件 $A_1, A_2, \cdots, A_N$ 的情形;也就是說,我們想要計算事件“至少有一个 A_k 发生”的概率,这事件可用符号 $A = A_1 \cup A_2 \cup \cdots \cup A_N$ 来表示. 为此,仅知道每个事件 A_k 的概率是不够的,我們还必須知道关于事件 A_k 的所有可能重迭的全部知識, 也就是对于每个二元組 (i, j), 每个三元組 (i, j, k),等等,我們必須知道 A_i 与 A_j, 或 A_i, A_j 与 A_k 等等同时发生的概率. 为了便于表达,我們将用字母 p 加以适当的下标来表示这些概率. 于是

$$p_i = \mathbf{P}\{A_i\},\ p_{i,j} = \mathbf{P}\{A_iA_j\},\ p_{i,j,k} = \mathbf{P}\{A_i\ A_j\ A_k\},\ \cdots \qquad (1.2)$$

---

1)此章內容在以后不很明显地用到,仅第一个定理相当重要.

此处,下标的先后次序是无关重要的. 但是,为了統一起見,我們将按下标数字的大小順序来书写;因此,我們不写 $p_{7,3,11}$ 而写 $p_{3,7,11}$. 二个下标永不相等. 令 S_r 表示所有具有 r 个下标的 p 之和, 即定义

$$S_1 = \sum p_i,\ S_2 = \sum p_{i,j},\ S_3 = \sum p_{i,j,k}, \cdots \tag{1.3}$$

其中 $i < j < k < \cdots \leqslant N$, 因此,在和里每一組合出現一次而且只是一次;故 S_r 共有 $\binom{N}{r}$ 个項.最后的一个和 S_N 仅有一項 $p_{1,2,3,\cdots,N}$, 它就是所有 N 个事件同时实現的概率. 当 $N = 2$ 时,只有两項 S_1 与 S_2, 并且公式(1.1)可写成

$$\mathbf{P}\{A\} = S_1 - S_2. \tag{1.4}$$

对于任意 N 个事件的情况,可推广成如下的定理.

定理. 在事件 $A_1, A_2, \cdots, A_N$ 中,至少有一个实現的概率 P_1 由下式給出:

$$P_1 = S_1 - S_2 + S_3 - S_4 + - \cdots \pm S_N. \tag{1.5}$$

証. 用所謂"包含与排斥"的方法来証明(1.5)(参看問題 26). 为了計算 P_1, 我們要把含在至少一个 A_i 中的所有样本点的概率都加起来,而每个这种点的概率只能加一次. 为了有系統地进行,首先累加那些只含在一个 A_i 中的点的概率,其次是那些恰含在两个事件 A_i 中的, 依此类推, 最后是那些含在所有 A_i 中的点(如果有的話)的概率. 現在, 設 E 为任一样本点, 它恰好含在这 N 个事件的 n 个中. 不失一般性,我們可以将事件組 $\{A_i\}$ 重新編号,使編号之后, E 恰好含在 $A_1, A_2, \cdots, A_n$ 中,而不在 $A_{n+1}, A_{n+2}, \cdots, A_N$ 中. 于是, $\mathbf{P}\{E\}$ 将出現在各个 $p_i, p_{i,j}, p_{i,j,k}, \cdots$ 之中,此处的下标数字都是从 1 到 n. 故 $\mathbf{P}\{E\}$ 在 S_1 中出現 n 次,在 S_2 中出現 $\binom{n}{2}$ 次等等. 总之,当(1.5)的右边用样本点的概率来表示时,得到了 $\mathbf{P}\{E\}$ 出現的次数是

$$n - \binom{n}{2} + \binom{n}{3} - + \cdots \pm \binom{n}{n}. \tag{1.6}$$

要証明这个定理,只須証明这个数等于 1 便够了. 把(1.6)与 $(1-1)^n$

的二項展开式[参看第二章(8.7)]作比較，立刻就可以得到这个結論．因为 $(1-1)^n$ 的展开式中的第一項是 1 ，而紧接着的各項便是(1.6)各項的反号．故对于每个 $n \geqslant 1$，表达式(1.6)都等于 1 ，这就証明了定理．

例． (a)在桥牌游戏中，設 A_i 为“第 i 家有同一花色”的事件，則 $p_i = 4/\binom{52}{13}$；第 i 与第 j 两家同时都有同一花色的事件的发生共有 4・3 种可能的方式，并且有概率 $p_{i,j} = 12/\binom{52}{13}\binom{39}{13}$；同理，

$$p_{i,j,k} = 24/\binom{52}{13}\binom{39}{13}\binom{26}{13}.$$

最后，$p_{1,2,3,4} = p_{1,2,3}$，因为当三家中每家都有同一花色时，第四家必然也有同一花色． 因此某一家有同一花色的概率是 $P_1 = 4p_1 - 6p_{1,2} + 4p_{1,2,3} - p_{1,2,3,4}$． 利用史特令公式，可以近似地得到 $P_1 = \frac{1}{4}10^{-10}$． 在这个特殊的情形，$P_1$ 很相近于 A_i 的概率之和，但这并不是常規，而是一个例外．

(b) 相合． 孟德模 (Montmort) (1708) 首先提出下面的有种种变形和一个奇妙的解答的問題．它曾被拉普拉斯与許多其他作者推广过．

設有两副同样的紙牌，每副各有 N 张不同的牌． 把这两副牌都均匀地洗过之后，就将此两副牌依它們的次序逐一翻看对比．若在两副牌中，同一位置的两张牌恰好是相同时，我們就說有一个相合．因而，相合可以在 N 个位置的任何一处发生，也可以在若干个位置上同时发生． 这样的試驗还可以用更有趣的形式来描写．例如，我們可用 N 封信与 N 个信封来代替两副紙牌，而設想一个馬虎的秘书把信与信封随意地相配． 或者，我們也可以把它設想如下：把混放在衣帽間的帽子随便地发还給宾客，若某一客人所得的帽子恰好是他自己的，則說有一个相合．我們值得来想一下一个相合发生的概率与 N 的关系：在一个有 8 个客人的餐館中，帽主与其帽子有一个相合的概率同 10,000 人中有一个相合的概率相比較将如何呢？人們将感到奇怪，实际上，这概率差不多与 N 无关，总是近乎 2/3 (对于較有价值的应用，可参看問題 10，11)．

在第 4 节里，我們将計算恰有 0，1，2，3，…个相合的概率．这里，我們仅仅求出至少有一个相合的概率 P_1．为了表达簡单起見，設我們将紙牌編上号数 1，2，3，…，N，幷假定有一副紙牌是按它自然的順序排列的，而另一副紙牌的每一种排列都有概率 $\frac{1}{N!}$．令 A_k 为在第 k 个位置上这两副紙牌出現相合的事件．这意味着：編号为 k 的牌排在第 k 个位置，而其余的 $N-1$ 张牌可以任意的順序来排列．显然，$p_k=\frac{(N-1)!}{N!}=\frac{1}{N}$．同理，对每一組 i,j，我們有 $p_{i,j}=\frac{(N-2)!}{N!}=\frac{1}{N(N-1)}$，等等．在和 S_r 中包含有 $\binom{N}{r}$ 个項，其中每一項都等于 $\frac{(N-r)!}{N!}$．因此 $S_r=\frac{1}{r!}$ 及由(1.5)即可得到所要求的概率是

$$P_1=1-\frac{1}{2!}+\frac{1}{3!}-+\cdots\pm\frac{1}{N!}. \tag{1.7}$$

注意，$1-P_1$ 正是展开式

$$e^{-1}=1-1+\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}-+\cdots \tag{1.8}$$

中的前 $N+1$ 項．因此我們有一个良好的近似式

$$P_1\approx 1-e^{-1}=0.63212\cdots. \tag{1.9}$$

下表中 P_1 的准确数值可說明(1.9)的近似程度是很好的．

$N=$	3	4	5	6	7
$P_1=$	0.66667	0.62500	0.63333	0.63196	0.63214

2．在古典占位問題中的应用

現在回到把 r 个球随机地放到 n 个盒中的問題，設每种不同的放法都具有相同的概率 n^{-r}．我們要求出恰有 m 个盒是空的概率 $p_m(r,n)$[1]．

設 A_k 代表第 k 个盒是空的这一事件($k=1,2,\cdots,n$)．在

1) 这个概率在第二章問題(11.8)中曾用完全不同的方式推出过．可把它与第 3 节中的例子比較．

这事件中,所有 r 个球都放到其余的 $n-1$ 个盒里,并且,这样的放法共有$(n-1)^r$ 种. 同理,預先指定的两个盒是空的放法共有$(n-2)^r$ 种,依此类推. 于是得到

$$p_i=\left(1-\frac{1}{n}\right)^r,\quad p_{i,j}=\left(1-\frac{2}{n}\right)^r,\quad p_{i,j,k}=\left(1-\frac{3}{n}\right)^r,\cdots,\tag{2.1}$$

因此,对于每个 $\nu\leqslant n$, 有

$$S_\nu=\binom{n}{\nu}\left(1-\frac{\nu}{n}\right)^r.\tag{2.2}$$

至少有一个盒是空的概率由(1.5)給出, 因此所有的盒都不空的概率为 $1-S_1+S_2-+\cdots$,或者

$$p_0(r,n)=\sum_{\nu=0}^{n}(-1)^\nu\binom{n}{\nu}\left(1-\frac{\nu}{n}\right)^r.\tag{2.3}$$

現在我們考慮一个恰有 m 个盒空的分布. 这 m 个盒可以有 $\binom{n}{m}$ 种方法选取. r 个球分布在其余的 $n-m$ 个盒中, 因此这样的分布共有$(n-m)^r p_0(r,n-m)$种. 除以 n^r, 我們发現恰有 m 个盒是空的概率为

$$p_m(r,n)=\binom{n}{m}\left(1-\frac{m}{n}\right)^r p_0(r,n-m)=$$

$$=\binom{n}{m}\sum_{\nu=0}^{n-m}(-1)^\nu\binom{n-m}{\nu}\left(1-\frac{m+\nu}{n}\right)^r.\tag{2.4}$$

我們曾經采用 r 个随机数字的模型来譬喻 r 个物件任意地放入 10 个盒中. 此时,空盒对应于数字的不出現. 若 m 个盒是空的,則 10—m 个不同数字就出現在所給出的序列中. 表 1 供給我們一个数值的实例.

显然, 在(2.4)中, 直接的数字計算是在 n 与 r 都較小时才是适用的, 但占位問題只在 n 相当大时才特别有兴趣. 例如, 若 10,000 个球任意地放到 1000 个盒中,究竟有多少机会发現一个空盒? 在人数为 2000 的人羣中,发現一年中的某一天不是这些人的生日的机会又是多少? 幸好, 这类問題可以利用 $P_{[m]}$ 的一个很簡

表 1. 当 $n=10$ 时,按照公式(2.4)所計算的概率 $p_m(r, 10)$

m	$r=10$	$r=18$
0	0.000 363	0.134 673
1	.016 330	.385 289
2	.136 080	.342 987
3	.355 622	.119 425
4	.345 144	.016 736
5	.128 596	.000 876
6	.017 189	.000 014
7	.000 672	.000 000
8	.000 005	.000 000
9	.000 000	.000 000

$p_m(r, 10)$ 是在一列 r 个随机数字中,0,1,2,…,9 中恰有 m 个不出現的概率.

洁而又令人满意的近似法来回答. 当 $n\to\infty$ 时,其誤差趋于零. 有关这近似法的討論是概率論中許多极限定理的典型.

于是,我們的目的在于討論当 $n\to\infty$, $r\to\infty$ 时公式(2.4)的极限形式. 在原則上,r 与 n 之間的关系是任意的. 然而,比例 $\frac{r}{n}$ 是表示每一盒中物件的平均数. 如果 $\frac{r}{n}$ 很大,我們就不能期望有空盒. 在这情形,$p_0(r, n)$ 近似于 1 而所有的 $p_m(r, n)(m\geqslant 1)$ 都很小. 另一方面,若 $\frac{r}{n}$ 趋于零,則几乎所有的盒都是空的. 此时,对每一固定的 m, $p_m(r, n)\to 0$. 因此,只有中間情形才眞正有討論的意义.

我們首先估計公式(2.2)中的量 S_ν. 因为 $(n-\nu)^\nu<(n)_\nu<n^\nu$, 显然,我們有

$$n^\nu\left(1-\frac{\nu}{n}\right)^{\nu+r}<\nu!\,S_\nu<n^\nu\left(1-\frac{\nu}{n}\right)^r. \tag{2.5}$$

对 $t=\nu/n$ 应用第二章(8.12)的两重不等式,我們得到

$$\{ne^{-(\nu+r)/(n-\nu)}\}^\nu<\nu!S_\nu<\{ne^{-r/n}\}^\nu. \tag{2.6}$$

为了簡单起見,令

$$ne^{-r/n}=\lambda, \tag{2.7}$$

并假定 r 和 n 以这样一种方式上升：使 λ 保持有界．于是，对每一固定的 ν，(2.6) 中最远的两端的数的比趋于 1，因此，我們推出

$$S_\nu < \frac{\lambda^\nu}{\nu!} \text{ 和 } \frac{1}{\nu!}\lambda^\nu - S_\nu \to 0^*. \tag{2.8}$$

由此推出

$$p_0(r, n) - \sum_{\nu=0}^{\infty}(-1)^\nu \frac{\lambda^\nu}{\nu!} \to 0, \tag{2.9}$$

或者 $p_0(r, n) - e^{-\lambda} \to 0$．但(2.4)中 $p_0(r, n-m)$的因子可以写成 S_m，从而，对每一个固定的 m，都有

$$p_m(r, n) - e^{-\lambda}\frac{\lambda^m}{m!} \to 0. \tag{2.10}$$

这就完成了下述定理的証明：

定理[1]．当 n 和 r 增大而使得 $\lambda = ne^{-\frac{r}{n}}$保持有界时，则对每一个固定的 m，(2.10)都成立．

近似表达式

---

* 在第一版的証明中，作者省去了这一步，而直接考虑 $\frac{r}{n^2} \to 0$ 的情況．在第二版中补充了这一步，但証明不完全确切．例如令 $r = n^2$ 则 $\lambda = ne^{-\frac{r}{n}} = ne^{-n} \to 0$，当 $n \to \infty$，这符合 λ 保持有界的条件，但(2.6)两边之比

$$\left\{e^{-\frac{\nu+r}{n-\nu}+\frac{r}{n}}\right\}^\nu = e^{-\frac{\nu^2(n+r)}{n(n-\nu)}} = e^{-\frac{\nu^2(n+n^2)}{n^2-n\nu}} \to e^{-\nu^2}$$

一般不趋于 1．故这一步的証明应略加改正及补充如下：

在題設的条件下，如果：

(i) $\lambda \to 0$，则由 (2.6) $S_\nu < \frac{\lambda^\nu}{\nu!}$ 推得 $\frac{1}{\nu!}\lambda^\nu - S_\nu \to 0$．

(ii) 如 $n \to \infty$ 时，$\lambda > k_1 > 0$，则因 λ 保持有界，故存在 k_2，$0 < k_1 < \lambda < k_2 < \infty$．这时 $\frac{k_1}{n} < e^{-\frac{r}{n}} < \frac{k_2}{n}$，因而 $\sqrt[n]{\frac{k_1}{n}} < e^{-\frac{r}{n^2}} < \sqrt[n]{\frac{k_2}{n}}$，故 $e^{-\frac{r}{n^2}} \to 1$，亦即 $\frac{r}{n^2} \to 0$．这时 (2.6) 两边之比为 $e^{-\frac{\nu^2(n+r)}{n(n-\nu)}} \to 1$．故仍有 $\frac{1}{\nu!}\lambda^\nu - S_\nu \to 0$．

一般情況(即对应于整数 n (或 r)，λ 有一子序列收斂于 0，另一子序列保持大于某一大于 0 的常数)，均能保証上面极限关系的成立——譯者注．

1) 馮・米賽斯給出了一个不同的証明方法，参看[28]．

表 2. 把 r 个球随机地放入 1000 个盒里,其中恰有 m 个空盒的概率的普阿松近似值(2.11)

r	λ	$p(m;\lambda)$											
		$m=0$	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$	$m=7$	$m=8$	$m=9$	$m=10$	$m=11$
5000	6.74	0.0012	0.0080	0.0269	0.0604	0.1017	0.1371	0.1540	0.1482	0.1249	0.0935	0.0630	0.0386
5500	4.09	.0167	.0685	.1400	.1909	.1951	.1596	.1088	.0636	.0325	.0148	.0060	.0023
6000	2.48	.0838	.2077	.2575	.2128	.1320	.0655	.0271	.0096	.0030	.0008	.0002	
6500	1.50	.2231	.3347	.2510	.1255	.0471	.0141	.0035	.0008	.0001			
7000	0.91	.4027	.3661	.1666	.0506	.0115	.0021	.0003					
7500	0.55	.5777	.3163	.0873	.0162	.0023	.0003						
8000	0.34	.7126	.2406	.0414	.0049	.0004							
8500	0.20	.8187	.1637	.0164	.0011	.0001							
9000	0.12	.8869	.1064	.0064	.0003								

$$p(m;\lambda)=e^{-\lambda}\frac{\lambda^m}{m!} \tag{2.11}$$

称为普阿松分布，它是一个非常重要的分布，它描述了各种現象；在第六章将要仔細研究它.

实际上，当 n 相当大时，就可用 $p(m,\lambda)$ 作为一个逼近. 而对于不大不小的值 n，就需要估計誤差. 但是在这里我們不从事于誤差估計的討論了.

例. (a) 当盒的个数是 1000 而球的个数在 5000 与 9000 之間时，表 2 給出了恰有 m 个空盒的概率的近似值. 当 $r=5000$ 时，空盒个数的中位数是 6：出現 7 个或更多的空盒的可能性与出現 6 个或更少的空盒的可能性是差不多的. 如果把 9000 个物件任意分配到 1000 个盒里，则可发現一个空盒的机会是 1/9.

(b) 在生日的統計中 [第二章例(3.d)]，$n=365$，r 是人数. 当 $r=1900$ 时，則近似地有 $\lambda=2$. 在一个 1900 人的乡村里，一年中有 m 个日子不是生日的概率可近似地表之如下：

$$P_{[0]}=0.135,\qquad P_{[1]}=0.271,$$
$$P_{[2]}=0.271,\qquad P_{[3]}=0.180,$$
$$P_{[4]}=0.090,\qquad P_{[5]}=0.036,$$
$$P_{[6]}=0.012,\qquad P_{[7]}=0.003.$$

用同样的方法可以推出：恰有 m 个盒，每个都含有 k 个物体的概率. 而馮·米賽斯曾經指出：这个概率也可以用普阿松表示式(2.11)来逼近，只不过要把 λ 定义为

$$\lambda=ne^{-r/n}\left(\frac{r}{n}\right)^k\Big/k!. \tag{2.12}$$

3. N 个事件中实現 m 件

第 1 节的定理可以加強如下面的

定理. 对任何正整数 m，$1\leqslant m\leqslant N$，在 N 个事件 $A_1,\cdots,A_N$ 中恰有 m 个同时发生的概率 $P_{[m]}$ 为

$$P_{[m]}=S_m-\binom{m+1}{m}S_{m+1}+\binom{m+2}{m}S_{m+2}-+$$

$$+\cdots\pm\binom{N}{m}S_N. \tag{3.1}$$

注意：根据(1.5)，沒有一个 A_j 发生的概率 $P_{[0]}$ 为

$$P_{[0]}=1-P_1=1-S_1+S_2-S_3\pm\cdots\mp S_N. \tag{3.2}$$

这指出了，如果取 $S_0=1$，(3.1)式对 $m=0$ 时也是合理的.

証. 我們采用(1.5)的証明方法. 設 E 是任意一个样本点，并設它恰好含在 N 个事件 A_j 的 n 个中，則仅当 $n=m$ 时样本点 E 的概率 $\mathbf{P}\{E\}$ 才是概率 $P_{[m]}$ 的組成部分. 現在，我們来观察 $\mathbf{P}\{E\}$ 在(3.1)右边出現的情况，注意，$\mathbf{P}\{E\}$ 只在和 $S_1, S_2, \cdots, S_n$ 中出現，而不在 $S_{n+1}, S_{n+2}, \cdots, S_N$ 中出現. 于是，当 $n<m$ 时，$\mathbf{P}\{E\}$ 在(3.1)右边不出現. 当 $n=m$ 时，$\mathbf{P}\{E\}$ 恰在 S_m 中的一項出現. 要完成定理的証明，只要証明下面的情况便可以了，即当 $n>m$ 时，在(3.1)右边的項 $S_m, S_{m+1}, \cdots, S_n$ 中，$\mathbf{P}\{E\}$ 的出現的次数应該恰好相消. 由于包含 E 的 n 个事件可以組成 $\binom{n}{k}$ 个 k 重組，故 $\mathbf{P}\{E\}$ 在 S_k 中共出現 $\binom{n}{k}$ 次. 因此对 $n>m$，$\mathbf{P}\{E\}$ 在(3.1)右边出現总的次数是

$$\binom{n}{m}-\binom{m+1}{m}\binom{n}{m+1}+\binom{m+2}{m}\binom{n}{m+2}-+\cdots\mp$$
$$\mp\binom{n}{m}\binom{n}{n}. \tag{3.3}$$

因为 $\binom{m+v}{m}\binom{n}{m+v}=\binom{n}{m}\binom{n-m}{v}$，所以(3.3)可写成

$$\binom{n}{m}\left\{\binom{n-m}{0}-\binom{n-m}{1}+\cdots\mp\binom{n-m}{n-m}\right\} \tag{3.4}$$

在花括号內恰好是 $(1-1)^{n-m}$ 的二項展开式，因此(3.3)式化为零. 这正如我們所断言的.

例. 讀者可以驗証：把(2.2)代入(3.1)直接推出公式(2.4).

4. 在相合与猜測問題中的应用

在例(1.b)中，我們考虑两副紙牌的相合問題，并求出 $S_k=1/k!$. 将它代入(3.1)就得到下面的結果：

設有两副同样的紙牌，每副各有N张不同的牌，在两副牌的随意配对中，恰有m个相合的概率$P_{[m]}$可以给出如下

$$
\begin{aligned}
P_{[0]}&=1-1+\frac{1}{2!}-\frac{1}{3!}+-\cdots\pm\frac{1}{(N-2)!}\mp\frac{1}{(N-1)!}\pm\frac{1}{N!},\\
P_{[1]}&=1-1+\frac{1}{2!}-\frac{1}{3!}+-\cdots\pm\frac{1}{(N-2)!}\mp\frac{1}{(N-1)!},\qquad (4.1)\\
P_{[2]}&=\frac{1}{2!}\left\{1-1+\frac{1}{2!}-\frac{1}{3!}+-\cdots\pm\frac{1}{(N-3)!}\mp\frac{1}{(N-2)!}\right\},\\
P_{[3]}&=\frac{1}{3!}\left\{1-1+\frac{1}{2!}-\frac{1}{3!}+-\cdots\pm\frac{1}{(N-3)!}\right\},\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
P_{[N-2]}&=\frac{1}{(N-2)!}\left\{1-1+\frac{1}{2!}\right\},\\
P_{[N-1]}&=\frac{1}{(N-1)!}\{1-1\}=0,\\
P_{[N]}&=\frac{1}{N!}.
\end{aligned}
$$

最后的一个关系是显然的. $P_{[N-1]}$为零表示：有$N-1$个相合而沒有所有的N个相合是不可能的事件.

在(4.1)式右边的花括号內恰是e^{-1}展开式中的前若干項. 故对于相当大的N，我們有近似式

$$P_{[m]}\approx\frac{1}{m!}e^{-1}. \qquad (4.2)$$

下頁表3中給出了$N=3, 4, 5, 6, 10$时所对应的諸$P_{[m]}$的准确值与极限值，每行中的前一行給出了$P_{[m]}$的准确值，而最后的一行却給出了它的极限值

$$p_m=\frac{e^{-1}}{m!}. \qquad (4.3)$$

从表中可以看到p_m对$P_{[m]}$的近似是相当良好的，甚至对于不大的N值也行.

(4.3)所定义的数值p_m有如下的性質：$\sum p_k=e^{-1}\Big(1+1+$

表 3. 对一付 N 张不同的牌进行猜测而恰有 m 次猜中的概率

	$N=3$		$N=4$		$N=5$		$N=6$		$N=10$		p_m
	$P_{[m]}$	b_m	$P_{[m]}$	b_m	$P_{[m]}$	b_m	$P_{[m]}$	b_m	$P_{[m]}$	b_m	
0	0.333	0.296	0.375	0.316	0.367	0.328	0.368	0.335	0.36788	0.34868	0.367879
1	.500	.444	.333	.422	.375	.410	.367	.402	.36788	.38742	.367879
2	···	.222	.250	.211	.167	.205	.187	.201	.18394	.19371	.183940
3	.167	.037	···	.047	.083	.051	.056	.053	.06131	.05740	.061313
4			.042	.004	···	.006	.021	.008	.01534	.01116	.015328
5					.008	.000	···	.001	.00306	.00149	.003066
6							.001	.000	.00052	.00014	.000511
7									.00007	.00001	.000073
8									.00001	···	.000009
9									···	···	.000001
10									···	···	.000000

$P_{[m]}$ 由(4.1)給出，b_m 由(4.4)給出．最后一行由普阿松极限(4.3)給出．

$\left.\frac{1}{2!}+\frac{1}{3!}+\cdots\right)=e^{-1}e=1$. 因此，$p_k$ 可以解释为概率．注意：(4.3)为普阿松分布(2.11)当 $\lambda=1$ 的特别情形．

公式(4.1)在試驗猜測的能力时是有用的．在尝酒鉴定及在心理試驗等等，是使猜測者說出(或呼喚)一組未知順序的 N 个物件，例如紙牌．要是猜的人眞有点"秘訣"的話，就会有非随机性表現出来．要想从他的猜中率来判断有沒有"秘訣"，我們須要算一算运气的概率．随机的猜测可依照几种方式来进行，而我們仅叙述其中三个极端的可能情形．1)猜测者始終保持呼喚一个固定的紙牌，则在 N 次呼喚中，他一定有一个且仅有一个猜中，因此，猜中次数沒有随机的起伏．2)在 N 次呼喚中，猜測者每张牌都呼喚一次，則 N 次呼喚就对应于 N 张紙牌的一个排列，若猜測者沒有什么秘訣，則公式(4.1)可以应用．3)在 N 次呼喚中，猜測者每次皆可猜測 N 张牌中的任意一张，即 N 次的猜測是彼此独立地进行．此时，共有 N^N 种可能的排列．实际上，每人有其固定的心理习惯，往往会傾向于呼喚某些牌比其他的牌要多一些，但在初步近似中，我們可以假設所有 N^N 种排列都是等概的．由于猜中 m 个而其余

$N-m$ 个不猜中的排列共有 $\binom{N}{m}(N-1)^{N-m}$ 种，故恰有 m 个猜中的概率是

$$b_m = \binom{N}{m}\frac{(N-1)^{N-m}}{N^N}. \tag{4.4}$$

(这是二項分布的特殊情形,在第二章例(4.c)中已推出.)

表 3 給出了按照方式 2)和方式 3)进行試驗时,其猜中概率的比較. 为了判断两种方式的短长，我們需要均値与随机起伏的理論. 在这两种方式下,猜中的平均次数都是 1 ,而方式 2)的随机起伏比方式 3)来得大些. 从表 3 中可看到,实际上,它們之間的差别不是太大.

5. 杂　　录

(a) 至少有 m 个事件实現. 沿用第 3 节的符号,在事件 A_1, A_2, $\cdots$, A_N 中,有 m 个或更多个事件同时发生的概率 P_m 可以由下式給出:

$$P_m = P_{[m]} + P_{[m+1]} + \cdots + P_{[N]}. \tag{5.1}$$

要求由 S_k 来表达 P_m 的公式,最簡易的途徑是采用归納法. 从公式(1.5) 出发，应用递推公式 $P_{m+1} = P_m - P_{[m]}$ 卽可得到，当 $m \geqslant 1$ 时,有

$$P_m = S_m - \binom{m}{m-1}S_{m+1} + \binom{m+1}{m-1}S_{m+2} - \binom{m+2}{m-1}S_{m+3} + \\ + \cdots \pm \binom{N-1}{m-1}S_N. \tag{5.2}$$

此外,还可用(3.1)的結果直接导出(5.2).

(b) 其他恒等式. 我們用 $P_{[k]}$ 或 P_k 来表达 S_ν, 玆叙述如下:

$$S_\nu = \sum_{k=\nu}^{N}\binom{k}{\nu}P_{[k]}, \tag{5.3}$$

$$S_\nu = \sum_{k=\nu}^{N}\binom{k-1}{\nu-1}P_k. \tag{5.4}$$

証明提示. 对于給定的値 $P_{[m]}$, 方程(3.1)可視为未知数 S_ν

的綫性方程，我們要証明(5.3)是它的唯一解．若把(5.3)代入 $P_{[m]}$ 的表达式(3.1)，則右边 $P_{[k]}$ $(m \leqslant k \leqslant N)$的系数是

$$\sum_{\nu=m}^{k}(-1)^{\nu-m}\binom{\nu}{m}\binom{k}{\nu}=\binom{k}{m}\sum_{\nu=m}^{k}(-1)^{\nu-m}\binom{k-m}{\nu-m}. \quad (5.5)$$

当 $k=m$ 时，表达式(5.5)等于 1．当 $k>m$ 时，(5.5)是$(1-1)^{k-m}$的二項展开式，故为零．因此，将(5.3)代入(3.1)后，即得恆等式 $P_{[m]}=P_{[m]}$．方程(3.1)的解的唯一性是根据这样的事实：每个 $P_{[m]}$ 的方程(3.1)只加一个新的未知数 S_m，故 S_ν 可以逐步地算出．我們可以用类似的方法証明(5.4)的正确性．

(c) 彭費雷尼(Bonferroni)不等式． 用下面的方法可以得出有关 $P_{[m]}$ 与 P_m 的一系列不等式．如果在(3.1)或(5.2)中，只保留 $S_m, S_{m+1}, \cdots, S_{m+r-1}$ 等項而把 $S_{m+r}, S_{m+r+1}, \cdots, S_N$ 等項抹去，則其誤差(即真值減去近似值)的符号与所抹去的第一項的符号(即$(-1)^r$)相同，且其絕对值是小于所抹去的第一項的絕对值．特別，当 $r=1$ 与 $r=2$ 时，有

$$S_m-(m+1)S_{m+1} \leqslant P_{[m]} \leqslant S_m, \quad (5.6)$$

$$S_m-mS_{m+1} \leqslant P_m \leqslant S_m. \quad (5.7)$$

証明提示． 要証明关于(3.1)的陈述，只須証明：对每一 t，有

$$\sum_{\nu=t}^{N}(-1)^{\nu-t}\binom{\nu}{m}S_\nu \geqslant 0. \quad (5.8)$$

利用(5.3)将不等式(5.8)的左边写成 $P_{[k]}$ 的綫性組合，則对于 $t \leqslant k \leqslant N$，$P_{[k]}$ 的系数等于

$$\sum_{\nu=t}^{k}(-1)^{\nu-t}\binom{\nu}{m}\binom{k}{\nu}=\binom{k}{m}\sum_{\nu=t}^{k}(-1)^{\nu-t}\binom{k-m}{\nu-m}.$$

后者之和为 $\binom{k-m-1}{t-m-1}$，所以 $P_{[k]}(t \leqslant k \leqslant N)$ 的系数是正的，故得証(参看第 2 章問題(12.13))．关于更深刻的不等式，讀者可参考本章开始时曾提到过的那篇弗雷謝的专著．

6. 問　　題

注． 假設在下面的每个情况，所有可能的排列都是等概的。

1．从放有 10 双鞋子的橱子里任取 4 只鞋子，試求所取出的 4 只鞋中至少有两只配对的概率。

2．一次擲出五个骰子，試求其中至少有三个点数相同的概率（并用第二章第 5 节的方法驗証之）。

3．把一个錢币扔 5 次，試求正面至少連續出現 3 次的概率。

4．把一个錢币扔 10 次，求正面至少連續出現 5 次的概率。

5．在上述两題中，以骰子代替錢币，么点代替正面，求其相应的概率。

6．把两顆骰子擲 r 次，試求在对子 $(1, 1), (2, 2), \cdots, (6, 6)$ 中，每对至少出現一次的概率 p_r。

7．一付桥牌中的«四同»． 四张面值相同的牌我們称之为四同． 因此，一付 13 张桥牌中可以有 0，1，2 或者 3 个四同．算出对应的概率．

8．有放回抽样． 从一个具有 n 个人的总体中抽取一个大小为 r 的样本． 求出 N 个指定的人全被抽在样本里的概率 u_r [这就是第二章問題 (11.12)]．

9．无放回抽样． 在現在无放回的情形回答类似問題 8 的問題，并証明問題 8 中有 $u_r \longrightarrow p^N$（这就是第二章問題(11.3)，不过現在的方法推出一个完全不同的公式）．

10．在 N 阶行列式的一般展开式中，包含一个或多个对角綫上的元素的項数共有 $N!P_1$，其中 P_1 由(1.7)所确定．

11．把 8 只城状的(国际象棋)棋子放在棋盘上，使得彼此都不能被吃掉而且沒有一个在白色的对角綫上．試証这种放法的总数是 $8!(1 - P_1)$，其中 P_1 是由(1.7)所定义，而 $N = 8$．

12．一个抽样（贈券收集者）問題． 一副由 s 类每类有 n 张所构成的牌中，每一类都标有数字 $1, 2, \cdots, n$．从这副牌无放回地抽取一个大小为 $r \geqslant n$ 的随机样本．算出每一个标号都出現在样本中的概率 u_r．（把这情形应用到 $s = 4, n = 13$ 的桥牌中去，我們得到一付 r 张桥牌中包含全部 13 种面值的概率；而应用到 $s = 13,\ n = 4$ 的情形中去，我們得到四种花色都拿到了的概率．）

13（續上）． 証明：当 $s \longrightarrow \infty$ 时有 $u_r \longrightarrow p_0(r, n)$，其中最后一个式子由(2.3)所定义．这就意味着： 在极限情形，我們的抽样变成了从一个

具有数字 1, 2, ⋯, n 的总体中的有放回的随机抽样。

14(續上). 从問題12的結果推出：如果 $r<n$，有

$$\sum_{k=0}^{n}(-1)^k\binom{n}{k}(ns-ks)_r=0,$$

当 $r=n$ 时有

$$\sum_{k=0}^{n}(-1)^k\binom{n}{k}(ns-ks)_n=s^n n!.$$

用求

$$\frac{1}{(1-x)^{ns-r+1}}\{1-(1-x)^s\}^n$$

在 $x=0$ 的 r 阶微商的办法驗証上述方程。

15. 在抽样問題 12 中，求出恰巧抽 r 次就得到了全部数字的概率。考虑 $s\longrightarrow\infty$ 的极限情形。

16. 一个細胞中含有 N 个染色体，其中任意两个之各部分都可以交换. 如果发生了 r 对交换(共有 $\binom{N}{2}^r$ 种不同的方式发生)，求出恰有 m 个染色体卷入其中的概率[1)].

17. 从一付紙牌中，任意取出 5 张求 5 张中恰好缺少 k 种花色的概率。

18. 試求在任取13张的桥牌中，恰有 k 对同花的爱司和老开的概率。

19. 复相合. 两副 N 张不同的牌，每副同时与第三副相同的牌进行随机相配，求恰有 m 个复相合的概率 u_m，并証明当 $N\longrightarrow\infty$ 时，$u_0\longrightarrow 1$ [这告訴我們当 $m\geqslant 1$ 时，$u_m\longrightarrow 0$]。

20. 复相合. 前題中的做法可修改如下：从 $2N$ 张牌中任意取出 N 张，并将此 N 张与第三副牌进行随机相配，求沒有一个相合的概率，并証明当 $N\longrightarrow\infty$ 时，这概率趋于 $\frac{1}{e}$.

21. 复相合. 在問題 20 中，若将两副牌改为 r 副牌，求其相应的概率。

22. 在古典占位問題中，恰有 m 个盒含有 k 个物体的概率 $p_{[m]}(k)$ 是

$$p_{[m]}(k)=\frac{(-1)^m n!r!}{m!n^r}\sum_j(-1)^j\frac{(n-j)^{r-jk}}{(j-m)!(n-j)!(r-jk)!(k!)^j},$$

式中的和号是对 $j\geqslant m$ 且 $j\leqslant n$，$kj\leqslant r$ 求和。

23. 証明当 $k=1$ 时，第 2 节最后一段的叙述是正确的。

24. 在波司-爱因斯坦的統計中，利用(3.1)可导出恰有 m 个空盒的概率。

25. 用第二章中的公式(11.14)来驗証 24 題中所得到的公式。

26. 对 N 作归納法証明公式(1.5)。

---

1) 当 $N=6$ 的情形，参看[29]。

第五章　条件概率．統計独立性

1. 条件概率

条件概率的概念是概率論中的一个基本工具，然而不幸的是，由于一个特殊粗陋的术語使它不是很簡明了．下面的討論是用一种比較自然的方法来导出其形式的定义．

預备例子． 設在N个人的总体中，有N_A个色盲者而且有N_H个女的．若事件A及H分别表示随机选取的一人是色盲的及是女的(随机选取的定义可参閱第2章第2节)，则

$$\mathbf{P}\{A\}=\frac{N_A}{N},\quad \mathbf{P}\{H\}=\frac{N_H}{N}. \tag{1.1}$$

現在，以所有女人組成的子总体代替总体的位置，我們来計算女人中随机选出的一人是色盲的概率．这概率是$\frac{N_{AH}}{N_H}$，其中N_{AH}是色盲女人的数目．这里，我們沒有用什么新的概念．但是，在研究某个特定的子总体时，我們需要用一个新的記号来表达．一般所采用的符号是$\mathbf{P}\{A|H\}$，可讀为“在事件H(所取出的人是女的)发生的假定下，事件A(色盲者)发生的概率”．采用符号

$$\mathbf{P}\{A|H\}=\frac{N_{AH}}{N_H}=\frac{\mathbf{P}\{AH\}}{\mathbf{P}\{H\}}. \tag{1.2}$$

显然，每一个子总体本身总可以被考虑为一个总体；为了語言上的方便起見，我們說一个子体时，意思就是說背后还有一个較大的总体．保险公司可以对由于雷击(事件A)引起的某种产值額的災害頻率是多少感覚兴趣．一般来說，保险公司的保险項目有好几类，如工业的、都市的、乡村的等等．单独研究工业項目的保险費意味着只研究事件A仅联系于事件H——“保险費是化在工业

項目的". 这时公式 (1.2) 显然可以应用. 然而注意: 对于一个只經营工业保险的保险公司来說, H 与整个样本空間重合, 因而 $\mathbf{P}\{A|H\}$ 变为 $\mathbf{P}\{A\}$.

最后,我們考虑坐在北边打桥牌的人. 一副桥牌分好了以后, 他知道了他手中的牌, 因而他感兴趣的是剩下的那 39 张牌的分布. 把这 39 张桥牌的一切可能的分布的集合作为样本空間是可以的,但是,把它們和北家手中的 13 张牌 (事件 H) 联系起来考虑并且討論假定事件 H 实現时事件 A(例如南家手中有二个爱司)的概率就会更方便一些. 公式(1.2)又能应用.

类似 (1.2) ,我們現在来引进下面正式的

定义. 令 H 是一个具有正概率的事件. 对于任何事件 A, 我們記

$$\mathbf{P}\{A|H\} = \frac{\mathbf{P}\{AH\}}{\mathbf{P}\{H\}}. \tag{1.3}$$

这样定义的量称为在 H 假設下 (或者給定 H) A 的条件概率. 当所有的样本点都是等概的时候, $\mathbf{P}\{A|H\}$ 就是 A 和 H 的公共的样本点的数目与 H 中的样本点的数目的比 N_{AH}/N_H.

当假設 H 的概率是零时, 条件概率 $\mathbf{P}\{A|H\}$ 就沒有定义了. 在离散样本空間的情形,这一点是无关重要的, 但是, 在一般理論中却是很重要的.

虽然記号 $\mathbf{P}\{A|H\}$ 本身是实用的, 但是它的詞句的表达却如此之长,以致实际上是采用較为簡略的描述, 例如, 在上面所引入的例中,我們用一个女人是色盲的概率来代替"在假定已知为女人的条件下, 随机选取一人是色盲者的条件概率". 我們經常以"若知道 H 已发生"来代替"在 H 的假定下". 总之, 我們的公式与記号是很明确的,而詞句的表达往往是非正式的,因而需要有正确的理解.

为了称呼上的清楚起見,对样本空間中的概率,有时我們用絕对概率这个名称. 严格地說, "絕对" 两字是多余的, 并且可以略去.

对于一个特定的假設 H，考虑各种事件的条件概率就相当于把 H 視为一个新的样本空間，并且，为了使新的样本空間的总概率为 1，故需将原空間中所有事件的概率都乘上因子 $1/\mathbf{P}\{H\}$. 这个叙述說明：所有概率的一般定理，对在任何特定假設 H 下的条件概率依然成立。例如，我們可以叙述 A 或 B 或二者同时发生的概率的基本关系如下：

$$\mathbf{P}\{A \cup B|H\} = \mathbf{P}\{A|H\} + \mathbf{P}\{B|H\} - \mathbf{P}\{AB|H\}. \quad (1.4)$$

类似，在第四章中，关于 N 个事件有 m 件实現的概率的全部定理改为条件概率的情形也仍然成立. 但是，我們并不需要它們.

我們往往应用以下面形式所表示的公式(1.3)

$$\mathbf{P}\{AH\} = \mathbf{P}\{A|H\}\mathbf{P}\{H\}. \quad (1.5)$$

这就是所謂复合概率定理. 为了推广到三个事件 A, B, C 的情形，可先取 $H = BC$ 作为假設事件，并再一次利用公式(1.5)，卽得

$$\mathbf{P}\{ABC\} = \mathbf{P}\{A|BC\}\mathbf{P}\{B|C\}\mathbf{P}\{C\}. \quad (1.6)$$

这可直接推广到四个或多个事件的情形.

最后，我們来討論一个簡单而又常用的公式. 設 $H_1, \cdots, H_n$ 是一組完备互斥事件（卽 $H_1, \cdots, H_n$ 之并集是整个样本空間，且 $H_1, \cdots, H_n$ 是互不相容的），則任何事件 A 只可能与某些 H_j 联合发生，卽

$$A = AH_1 \cup AH_2 \cup \cdots \cup AH_n. \quad (1.7)$$

由于諸 AH_j 是互不相容的，故 A 的概率是 AH_j 的概率之和，再利用 (1.5) 于 $H = H_j$，卽得

$$\mathbf{P}\{A\} = \Sigma \mathbf{P}\{A|H_j\}\mathbf{P}\{H_j\}. \quad (1.8)$$

这个公式是很有用的，因为計算条件概率 $\mathbf{P}\{A|H_j\}$ 有时要比直接計算概率 $\mathbf{P}\{A\}$ 容易一些.

例. (a) 无放回抽样. 从一个拥有 n 个元素 $1, 2, \cdots, n$ 的总体中抽取一个有序的样本. 令 i 和 j 是两个不同的元素，假定 i 是第一个被抽出来的元素（事件 H），第二个元素是 j（事件 A）的概率是多少？显然，$\mathbf{P}\{AH\} = \dfrac{1}{n(n-1)}$，$\mathbf{P}\{A|H\} = 1/(n-1)$.

这說明：第二次抽取是对于具有 $n-1$ 个元素的总体来进行的，其中每一个元素被抽中的概率都有相同的概率．事实上，随机抽样最自然的定义是："当前 r 次抽取完成后，剩下的 $n-r$ 个元素中的任何一个在第 $r+1$ 次被抽出的概率都是 $1/(n-r)$"．这个定义等价于第二章所給出的定义，但是我們不能在較早时叙述它，因为其中包含了条件概率的概念．

(b) 四个球連續地放进四个盒里，全部 4^4 个排列都是等可能的．給定前两个球放在不同的盒中（事件 H），某一个盒恰巧有三个球（事件 A）的概率为何？給定 H，事件 A 有两种不同的方式发生，所以 $\mathbf{P}\{A|H\}=2\cdot 4^{-2}=\frac{1}{8}$．（可以容易地直接驗証：事件 A 和 AH 分别包含 $12\cdot 4^2$ 和 $12\cdot 2$ 个样本点．）

(c) 性别的分布．考虑恰有两个小孩的家庭．今以 b, g 分别表示男孩或女孩．假定年长的孩子的性别用前一个字母来表示，則有四种可能的情形 bb, bg, gb, gg．这四个样本点的每一点都賦以概率 1/4．若已知一家有一个男孩（事件 H），問这一家的两个小孩都是男孩的概率是多少？因事件 AH 即 bb，而事件 H 为 bb 或 bg，或 gb，所以 $\mathbf{P}\{A|H\}=\frac{1}{3}$，即在具有特征 H 的所有家庭中，大約有 1/3 的家庭可以預期发生事件 A．这是很有趣味的，因大多数人所想象的解答总是 1/2．事实上，1/2 却是另一个不同問題的正确解答，这問題是：随机地遇到一男孩，并发現他是属于有两个小孩的家庭；問这家庭的另一个小孩也是男孩的概率为何？这两个問題的不同之处可解释如下：在第一个問題中，我們所面临的是一系列的家庭，而第二問題則是面临着一系列的男孩，在后者，有两个男孩的家庭共有两种可能情况，这就說明了这两个結果的不同之处．

(d) 分层总体．假定一羣人由一些子羣或者阶层 $H_1, H_2, \cdots$ 所构成．这些可以是民族，年龄，职业等等．令 p_j 为任取一人他属于 H_j 的概率．說"H_j 中的一个人是习惯用左手的概率是 q_j"，意思是指："在此人属于 H_j 的假定下事件 A（习惯用左手）的条件

概率”. 随机地抽取一个人,他是习惯用左手的概率为 $p_1q_1 + p_2q_2 + p_3q_3 + \cdots$, 这是 (1.8) 的一个特殊情形. 假定此人是习惯用左手的,他属于 H_j 的条件概率为

$$\mathbf{P}\{H_j|A\} = \frac{p_jq_j}{p_1q_1 + p_2q_2 + \cdots}. \tag{1.9}$$

2. 用条件概率所定义的概率.罐子模型

在前一节里,我們認为概率当然是定义在样本空間中的,并且只算了一些条件概率. 在应用中,許多实驗都是用某些特定的条件概率(虽然通常把形容詞“条件”省略了)来描述. 在理論上这意味着: 样本空間中的概率可由給定的条件概率中推导出来. 曾經指出过 (例 (1.a)) 无放回抽样的最好的定义是: 不管前 r 次抽取的結果是什么, 在第 $r+1$ 次抽取时, 剩下的每一个元素被抽出来的概率都是一样的. 类似地,在例子 (1. d) 中,我們的分层总体完全被几层的絕对概率 p_j 和每一层中的特性“习惯用左手”的条件概率 q_j 所描述. 用較多的例子来启示一般的模型比直接的描述更有效.

例. (a) 在第一章例 (5. b) 中,我們曾經考虑过三个游戏者 a, b, c 在一場游戏中互相替換的問題; 我們曾經描述过样本空間中的点子, 但是沒有賦概. 現在我們假定这一場游戏这样进行: 在每一次試驗中, 任何两个对手获胜的概率是 1/2. 虽然这种叙述并未包含“条件概率”这个詞, 但是却无形中与它有关联. 譬如說: 如果某游戏者 a 参与第 r 輪 (事件 H), 則他在这一輪获胜的概率是 1/2. 由方程(1.5)推出, 游戏者 a 在第一次和第二次游戏都获胜的概率是 1/4,用符号来表示,即是 $\mathbf{P}\{aa\} = \frac{1}{4}$. 重复地应用(1.5)可証: $\mathbf{P}\{acc\} = \frac{1}{8}$, $\mathbf{P}\{acbb\} = \frac{1}{16}$, 等等; 即是, 方案 (*) 中包含 r 个字母的样本点具有概率 2^{-r}. 这就是第一章問題 5 中所賦予的概率, 但是現在的描述更直观(在問題 14 中繼續).

(b) 家庭. 我們要解释下面的叙述. “一个家庭恰有 k 个小

孩的概率是 p_k（其中 $p_0 + p_1 + \cdots = 1$）. 对于每一个家庭来説，所有的性别的分布都是等概的”. 令 b 表示男孩 g 表示女孩，我們的样本空間由样本点 0（沒有小孩），b，g，bb，bg，gb，gg，bbb，$\cdots$ 所組成. 引号里面的第二个假定可以更明确地叙述为：如果知道該家庭恰有 n 个孩子，2^n 个可能的性别分布中的每一个都具有条件概率 2^{-n}. 而这假設的概率是 p_n，于是从 (1.5) 我們看出：n 个字母 b 和 g 的每一种排列的絕对概率都是 $p_n \cdot 2^{-n}$.

注意，这是分层总体的一个例子. 孩子数为 i 的家庭构成第 i 层 H_i. 作为一个例子，令 A 表示事件“該家庭有男孩无女孩”. 显然，其概率为 $\mathbf{P}\{A\} = p_1 2^{-1} + p_2 2^{-2} + \cdots$，这是 (1.8) 的一个特殊情形. 在这种情形下，假設 H_i 就是“家庭有 i 个小孩”. 現在我們問：如果知道一个家庭沒有女孩，該家庭仅有一个小孩的(条件)概率是多少？其中 A 是假設. 令 H 表示事件“仅有一个小孩”，則 AH 表示“一个小孩而沒有女孩”，而且

$$\mathbf{P}\{H \mid A\} = \frac{\mathbf{P}\{AH\}}{\mathbf{P}\{A\}} = \frac{p_1 2^{-1}}{p_1 2^{-1} + p_2 2^{-2} + p_3 2^{-3} + \cdots}, \quad (2.1)$$

这是 (1.9) 的一个特殊情形.

(c) **有后效的罐子模型.** 为了确定起見，我們考虑一个容易出事故的工业系統. 事故的发生可以用一个机会游戏的結果来描述. 一个罐子里有黑球和紅球，在每一段規定的时間区間里随机地取一个球，紅球代表事故出現. 如果事故出現的可能性对时間来說保持不变，則“罐子的成分”是常常相同的. 但是可以想象，每一个事件是有一个后效，使得或者增加或者減少一次新的事故发生的可能性. 这对应于一个成分可以改变的罐子，其成分根据依賴于順續抽取的結果的某种規則而改变. 构造出这样一个規則去概括各种情形是容易的，但是，我們只討論下面的模型[1]：

---

1) 用罐子模型去描述后效（传染病）似乎是由波利亚 (Polya) 提出的. 在一些文献中討論的許多模型都把他的方案（最初在[30]中引进）作为典范. 本书中所描述的模型及其三种特殊情形是由弗雷德曼 (B. Friedman)（参看[31]）所提出.

罐子模型．一个罐子中包含 b 个黑球与 r 个紅球．随机地抽取一个球．有放回地抽样，并且再加进 c 个与所抽出的球具有同样颜色的球和 d 个相反的颜色的球（这时罐子中就有 $r+b+c+d$ 个球了），这种手續反复地进行，其中 c 和 d 是任意的整数．c 和 d 可以取为負数，不过在这种情形下經过有限次取球以后会因为无球而停止．特别地，取 $c=-1, d=0$，則我們的抽样就变成了无放回的抽样，它在 $r+b$ 次以后就結束．

現在，我們轉向数学描述．注意：某些基本的概率可以通过它所确定的条件概率来計算．对应于 n 次抽取的样本空間的典型的描述法是用 n 个字母 B 和 R 的序列来代表其样本点．事件"第一次取出的是黑球"(即是第一个字母是 B 的全部序列所构成的集合）的概率为 $b/(b+r)$．如果第一个球是黑的，則第二次抽出的球仍为黑（条件）的概率为 $(b+c)/(b+r+c+d)$．因此，由(1.5)得知序列黑，黑(即是以 BB 开始的样本点的全体）的(絕对）概率为

$$\frac{b}{b+r}\,\frac{b+c}{b+r+c+d}. \tag{2.2}$$

序列黑，黑，黑的概率为(2.2)乘以 $(b+2c)/(b+r+2c+2d)$，等等．显然，用这种办法可以算出每一个样本点的概率．（当然，当 c 和 d 是負时，取球的次数 n 必須选得足够的小，以避免球的个数出現負数．）用归納法可以容易地驗証：所有的样本点的概率的和必須为 1．

概率的明显表达式不是很容易得到的，除非在下面要介紹的一个最重要的而且著名的特殊情形：

波利亚罐子模型，其特征是 $d=0, c>0$．每一次抽取以后，这时与取出的球有相同顏色的球的数目增加，而与取出的球的顏色不同的球的数目保持不变．在效果上看，每一次取出的球是什么顏色增加了下一次也取到这种顏色球的概率，因此，我們得到了一个如传染病現象的粗糙的模型，其中，每一次传染以后都增加再传染的概率．波利亚模型在分析上的簡单性是由下述明显的性質

所导出的：在 n 次抽取中，其結果为 n_1 个黑球 n_2 个紅球 $(n_1+n_2=$
$=n)$ 的任何一个序列的概率都与最初抽出 n_1 个黑球而后取出 n_2
个紅球的概率相同，卽是

$$p_{n_1,n}=$$
$$=\frac{b(b+c)(b+2c)\cdots(b+n_1c-c)\cdot r(r+c)\cdots(r+n_2c-c)}{(b+r)(b+r+c)(b+r+2c)\cdots(b+r+nc-c)}. \tag{2.3}$$

把分子分母都除以 c^n 并引用第二章(2.1)的符号，則上式可写为如下的形式：

$$p_{n_1,n}=\frac{\left(\frac{b}{c}+n_1-1\right)_{n_1}\left(\frac{r}{c}+n_2-1\right)_{n_2}}{\left(\frac{b+r}{c}+n-1\right)_n}$$
$$=\frac{\left(\frac{-b}{c}\right)_{n_1}\left(-\frac{r}{c}\right)_{n_2}}{\left(-\frac{b+r}{c}\right)_n}. \tag{2.4}$$

(波利亚模型在問題 18—24 中討論.)

除了波利亚模型以外，我們的罐子模型还包含另一个有趣的特殊情形，卽是：两个隔离壁之間热交換的爱伦弗斯特(Ehrenfest)模型[32,33]. 如物理学家所应用的那样，在最初的描述中，爱伦弗斯特模型考虑的是 k 个质点分布于两个容器中的情形. 随机地选取一个质点把它由它所在的容器移到另一个容器中去. 这种手續反复地进行. 間 n 步以后质点的分布如何? 为了把它化成罐子模型，只需把第一个容器中的质点叫作紅球，第二个容器中的质点叫作黑球. 于是在每一次抽取中，球抽出来換上一个相反的顏色的球，卽是 $c=-1, d=1$. 显然，在这种情形下，这种过程我們要継續多久就可以継續多久(如果沒有紅球，則自动地抽出一个黑球并換进一个紅球). [在第十五章例 (2. f) 中，我們将用另一种方法討論爱伦弗斯特模型.]

对 $c=0, d>0$ 的特殊情形，曾經由弗雷德曼提出作为一个

安全运动模型．每当发生了事故（即紅球被取出），安全运动就抓紧一些；而当沒有事故发生时，安全运动就放松一些，于是发生事故的概率就增大．

（d）层迭的罐子模型，假传染． 为了継續沿着上一个例子的綫索进行討論，我們假定每一个人都可能遇到意外事故，而且它們的发生取决于从一个罐子中随机地取一个球所得的結果． 然而，这时我們假定是无后效的，所以，罐子中的成份在整个过程中保持不变．因为意外事故出現的机会或傾向性可能会因人之不同而不同，会因职业之不同而不同，所以我們設想每一个人（或者每一种职业）都有他自己的罐子．为了避免不必要的复杂性，我們假定只有两种人（或两种职业），而且其数目之比为 1:5． 然后，我們考虑第一个罐子含有 r_1 个紅球与 b_1 个黑球，第二个罐子含有 r_2 个紅球与 b_2 个黑球． 实験"随机地选取一个人、并考察他在 n 个单位时間內发生了多少次意外事故"具有下述的对应关系：擲一顆骰子，如果么点出現，則选第一罐；否則选第二罐． 在每一种情形下，都是从这罐子中作 n 次有放回的随机抽取．我們的实験描述了一个保险公司接受一个新的保险戶的情形．

利用（1.8），我們得出：第一次抽出的球是紅的概率为

$$\mathbf{P}\{R\} = \frac{1}{6}\frac{r_1}{b_1 + r_1} + \frac{5}{6}\frac{r_2}{b_2 + r_2}, \tag{2.5}$$

而序列紅，紅（RR）的概率为

$$\mathbf{P}\{RR\} = \frac{1}{6}\left(\frac{r_1}{b_1 + r_1}\right)^2 + \frac{5}{6}\left(\frac{r_2}{b_2 + r_2}\right)^2. \tag{2.6}$$

在我們的模型里面并沒有包含什么数学問題，但是，它却具有一个很有趣的性質，这种性質在应用中会引起很大的混乱． 假定我們的保险公司观察出：一个新的保险戶在第一年中发生了一件意外事故，从而他关心这一个新的保险戶在第二年发生意外事故的概率．換句話說，假定第一次抽出的是紅球，我們間序列 RR 的（条件）概率是多少？显然，比 $\mathbf{P}\{RR\}/\mathbf{P}\{R\}$ 不等于$\mathbf{P}\{R\}$．为了解释这一点，我們假定 $r_1/(b_1 + r_1) = 0{,}6$ 和 $r_2/(b_2 + r_2) = 0{,}06$，

于是在任一次抽取中，取出紅球的概率为 0.15，但是，如果第一次抽出的是紅球，則第二次抽出的仍然是紅球的机会为 0.42．注意：我們的模型在全部的总体中是无后效的，但对一个随机地选出的人来說，他发生了一件意外事故以后会增加此人发生第二次意外事故的可能．这里，我們得到了抽样的影响；一次意外事故的发生幷不眞的产生什么影响，但是它却指出随机地选出的人发生意外的可能性有增大的傾向．

在統計文献中，习慣于用传染这个詞来代替后效．抽样的表面上后效最初誤解为眞实的传染效果，所以統計学家現在說到传染（或传染概率分布）时都是含糊不清的．例如，生态学家在田野里找寻昆虫．如果在一段无成效时間以后他找到了一个昆虫，他会断言：有巢穴，大概在附近，而且他再找到一些昆虫的机会是很大的．显然，这其中幷未包含任何后效性，而統計学家把这个說成是传染．

(e) 下面的例子是很著名的，同时具有例証的作用，只可惜有些人为造作．設想一个总体中有 $N+1$ 个罐子，每个罐子都有 N 个球，在第 k 号罐子中有 k 个紅球与 $N-k$ 个白球 $(k=0,1,2,\cdots,N)$．今从 $N+1$ 个罐中任意取出一罐，幷从这罐中作有放回的 n 次抽取．假設取出的 n 个球都是紅球（事件 A）．求在下一次（第 $n+1$ 次）仍然取出紅球（事件 B）的（条件）概率．設第一次所取的罐子是第 k 号，則在以后 n 次取球中都出現紅球的概率是 $\left(\frac{k}{N}\right)^n$．因此，由(1.8)我們有

$$\mathbf{P}\{A\}=\frac{1^n+2^n+\cdots+N^n}{N^n(N+1)}. \tag{2.7}$$

事件 AB 表示相継 $(n+1)$ 次取球中都出現紅球，于是有

$$\mathbf{P}\{AB\}=\mathbf{P}\{B\}=\frac{1^{n+1}+2^{n+1}+\cdots+N^{n+1}}{N^{n+1}(N+1)}. \tag{2.8}$$

故所求的概率是 $\mathbf{P}\{B|A\}=\mathbf{P}\{B\}/\mathbf{P}\{A\}$．

和(2.7)与(2.8)可考虑为近似于积分的黎曼和．因此，当 N 相

当大时，有

$$N^{-1}\sum_{k=1}^{N}\left(\frac{k}{N}\right)^n \sim \int_0^1 x^n dx = \frac{1}{n+1}. \tag{2.9}$$

故对足够大的N，我們有

$$\mathbf{P}\{B \mid A\} \approx \frac{n+1}{n+2}. \tag{2.10}$$

公式(2.10)可粗略地解释如下：若罐中所有的成分都是等概的，且在 n 次有放回的取球中都出現紅球，則在第 $n+1$ 次取球中仍然出現紅球的概率是 $(n+1)/(n+2)$. 这就是所謂拉普拉斯(1812)延續律.

关于貝依斯 (Bayes) 公式的附註. 在(1.9)与(2.2)中，我們曾經直接从定义出发来計算某些条件概率. 我們認为：初学者最好常常这样做，而不要去死記我們現在将要导出的公式(2.12). 公式(2.12)追溯到我們曾討过的特例的普遍情形，其实它仅是公式(1.3)的另一种写法. 設有一組完备互斥事件 $H_1, H_2, \cdots$，即每一个样本点必属于一个而且也只能属于一个 H_j，則我們所感兴趣的是

$$\mathbf{P}\{H_k \mid A\} = \frac{\mathbf{P}\{AH_k\}}{\mathbf{P}\{A\}}. \tag{2.11}$$

若将公式(1.5)与(1.8)代入(2.11)，則有

$$\mathbf{P}\{H_k \mid A\} = \frac{\mathbf{P}\{A \mid H_k\}\mathbf{P}\{H_k\}}{\sum_j \mathbf{P}\{A \mid H_j\}\mathbf{P}\{H_j\}}. \tag{2.12}$$

若将事件 H_k 称为原因，則(2.12)成为«关于原因概率的貝依斯法則». 在数学上，(2.12)不过是公式(1.3)的一种特殊写法而已，而在上述例(b)与(d)所描述的类型的許多統計应用中，公式(2.12)是很有用的. 但遺憾的是，由于例(e)类型的結果形式地乱用，以致使 Bayes 公式不被人信任. 在照例应用时，这种論証可能出毛病. 一个質量控制的工程师常是关心某架指定的机器而不管这架机器是否从无穷多个机器的总体中任意抽取出来的. 仅是基于邏輯上的可以接受同时又附合我們的想法，他就采用 Bayes 公式. Plato 曾用这种論証来証明大西洋的存在，而哲学家們用它来証明牛頓力学的荒謬. 但是对于我們工程师来說这种論証忽略了如下的情况：工程师期望成功，并且他可以估計与減小来自預測和猜測的各种錯誤，以便作得更滿意. 近代

統計检驗与統計估計的方法虽然較少直观性，但頗現实．因此，它不仅是被証明为正确的，且事实上可以应用．

3. 統計独立性

在上面各例中，条件概率 $\mathbf{P}\{A|H\}$ 一般地不等于絕对概率 $\mathbf{P}\{A\}$．通俗地說，我們对事件 H 是否发生的知識会改变事件 A 出現的把握．只有当 $\mathbf{P}\{A|H\}=\mathbf{P}\{A\}$ 时，这种知識才不能影响我們对于事件 A 的发生与否的推想．这时，我們就說：A 与 H 是統計独立的．但公式(1.5)表明了条件 $\mathbf{P}\{A|H\}=\mathbf{P}\{A\}$ 可以写成如下的形式：

$$\mathbf{P}\{AH\}=\mathbf{P}\{A\}\mathbf{P}\{H\}. \tag{3.1}$$

这等式关于 A 与 H 是对称的，并且还說明当 A 与 H 統計独立，H 与 A 也統計独立．因此，我們最好还从下面对称的形式来下定义：

定义 1. 两个事件 A 与 H 称为是統計独立的（或簡称独立），若等式(3.1)成立．当 $\mathbf{P}\{H\}=0$ 时，虽然 $\mathbf{P}\{A|H\}$ 是沒有意义的，但此定义仍可采用．統計独立是随机独立的同义詞．

例．（a）从一副紙牌中任取一张，由于对称的理由，我們可期望出現"黑桃"的事件与出現爱司的事件是独立的．事实上，它們的概率分别是 1/4，1/13；而它們同时发生的概率确是 1/52．

（b）抛擲两顆骰子，"第一顆骰子出現么点"的事件与"第二顆骰子出双数点"的事件是独立的．因为这两事件同时发生的概率是 $3/36=1/12$，而这正是它們分别的概率 1/6 与 1/2 之积．

（c）在 (a, b, c, d) 四个字母的排列中，"a 在 b 前"与"c 在 d 前"是独立的．这在直观上是显然的，并且也容易驗証．

（d）性别分布．我們回到例(1.c)，但現在所考虑的是有三个小孩的家庭，并假設所有八种可能情况 $bbb, bbg, \cdots, ggg$ 各有概率 1/8．令 H 是"家里男女孩子都有"的事件，而 A 是"家里至多有一个女孩"的事件．于是 $\mathbf{P}\{H\}=6/8$，而 $\mathbf{P}\{AH\}=3/8=\mathbf{P}\{A\}\mathbf{P}\{H\}$．于是，在家庭中有三个小孩的情况下，这两个事件是独立的．但是，当所考虑的家庭有两个或是四个小孩时，A 与 H 就不再独立．这說明了是否具有独立性并不总是显然的．

若事件H发生，則其逆事件H'就不发生，反之亦然．統計独立性隐含着不能从事件H的发生作出对事件A的任何推断．因此，A与H的独立性也就是A与H'的独立性（由于对称性，A'与H及A'与H'也是独立）．利用$\mathbf{P}\{H'\} = 1 - \mathbf{P}\{H\}$就很容易驗証这个結論．若公式(3.1)成立，則（因$AH' = A - AH$）

$$\mathbf{P}\{AH'\} = \mathbf{P}\{A\} - \mathbf{P}\{AH\} = \mathbf{P}\{A\} - \mathbf{P}\{A\}\mathbf{P}\{H\} = \\ = \mathbf{P}\{A\}\mathbf{P}\{H'\}, \tag{3.2}$$

这正是我們所期望的．

現在，設三个事件A, B与C是两两独立．于是

$$\begin{aligned} \mathbf{P}\{AB\} &= \mathbf{P}\{A\}\mathbf{P}\{B\}, \\ \mathbf{P}\{AC\} &= \mathbf{P}\{A\}\mathbf{P}\{C\}, \\ \mathbf{P}\{BC\} &= \mathbf{P}\{B\}\mathbf{P}\{C\}. \end{aligned} \tag{3.3}$$

人們可能猜想，两两独立总会隐含着象AB与C那样的一对事件的独立性．但是，这不一定对，我們可以举出反例說明(3.3)虽然成立而要事件A, B, C同时发生是不可能的，于是，AB与C就不会独立了．

例． (e) 擲两顆骰子，并定义如下的三个事件：A是"第一顆骰子出奇数点"，B是"第二顆骰子出奇数点"，最后C是"两顆骰子的点数和是奇数"（即一个是奇数点而另一个是偶数点）．若36个样本点的每一个都有概率1/36，則任何两个事件显然是独立的，且每一个事件的概率都是1/2．当已知其余两事件之一发生时，它的条件概率也是1/2．但三个事件不可能同时发生．当A发生而B不发生时則保証了C的必然发生．对于所有其他的組合也有类似的結果．

要使随机独立性不会有上述的情形发生，除(3.3)必需满足外，还要添上下面的假設：

$$\mathbf{P}\{ABC\} = \mathbf{P}\{A\}\mathbf{P}\{B\}\mathbf{P}\{C\}. \tag{3.4}$$

这等式保証了A与BC的独立性，并且也保証了B与AC，C与AB独立．甚至还可以証明$A \cup B$与C也是独立的．事实上，由第一章的基本关系式(7.4)，我們有

$$\mathbf{P}\{(A \cup B)C\} = \mathbf{P}\{AC\} + \mathbf{P}\{BC\} - \mathbf{P}\{ABC\}. \tag{3.5}$$

把(3.3)与(3.4)代入(3.5)的右边,則可提出一个公共因子 $\mathbf{P}\{C\}$,而余下的因子是 $\mathbf{P}\{A\} + \mathbf{P}\{B\} - \mathbf{P}\{AB\} = \mathbf{P}\{A \cup B\}$,于是

$$\mathbf{P}\{(A \cup B)C\} = \mathbf{P}\{A \cup B\}\mathbf{P}\{C\}. \tag{3.6}$$

这样一来,可以相信,在两个条件(3.3)和(3.4)成立之下,可以毫无困难地証明:凡能用 A 和 B 的組合所表示的事件均与 C 独立.

对于 n 个事件的普遍情形,我們有下面的定义:

定义 2. 事件 $A_1, A_2, \cdots, A_n$ 称为相互独立的,如果对所有可能的組合 $1 \leqslant i < j < k < \cdots \leqslant n$,以下的乘法規律成立:

$$\begin{aligned} &\mathbf{P}\{A_iA_j\} = \mathbf{P}\{A_i\}\mathbf{P}\{A_j\}, \\ &\mathbf{P}\{A_iA_jA_k\} = \mathbf{P}\{A_i\}\mathbf{P}\{A_j\}\mathbf{P}\{A_k\}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\mathbf{P}\{A_1A_2\cdots A_n\} = \mathbf{P}\{A_1\}\mathbf{P}\{A_2\}\cdots\mathbf{P}\{A_n\}. \end{aligned} \tag{3.7}$$

第一行代表$\binom{n}{2}$个等式,而第二行代表$\binom{n}{3}$个等式,…,等等.因此,所要满足的条件共有$\binom{n}{2} + \binom{n}{3} + \cdots \binom{n}{n} = (1+1)^n - \binom{n}{1} - \binom{n}{0} = 2^n - n - 1$个. 另一方面,第一行的$\binom{n}{2}$个条件足以保証这些 A_i 的两两独立性. 方程組(3.7)看起来似乎是一个复杂的条件組. 但是,它們的成立往往是很明显的,而且不需要驗証. 由归納法可知[从 $n = 2$ 和(3.2)出发]:

在定义 2 中,系統(3.7)可以由 2^n 个方程所构成的系統来代替,这些方程組是从(3.7)的最后一个方程把任意个 A_j 换为它們的互补事件 A_j' 而得出的.

两两独立与相互独立的区别在理論上的意义远胜于实际上的意义. 两两独立而不彼此独立的实际例子几乎是不存在的. 这种情形发生的可能性是由 S. 伯因斯坦发現的.

4. 重复試驗

統計独立性的概念使我們有可能分析地叙述"在同一条件下

重复"試驗的直观概念.

考虑代表臆想試驗的样本空間 $\mathfrak{S}$, 它的样本点是 $E_1, E_2 \cdots$, 并且它們的概率分别以 $p_1, p_2, \cdots$ 表示. 在两个相似試驗的体系中,所有可能結果 (E_j, E_k) 是組成一个新的样本空間. 对这些新样本点 (E_j, E_k), 概率的确定有很多种方式. 然而, 若实驗者假定: 两次試驗都是在同样条件下进行的, 卽隐含着独立性. 也就是說: 第一次試驗的結果不影响第二次的結果, 卽"第一次出現 E_j"的事件与"第二次出現 E_k"的事件是統計独立的,故有

$$\mathbf{P}\{E_j, E_k\} = p_j p_k. \tag{4.1}$$

对于每一可能結果 (E_j, E_k), 等式(4.1)給出一个相应的概率. 但在采用(4.1)作为新样本空間的概率定义以前, 我們必須証明这些 $p_j p_k$ 之和为 1. 在和 $\sum\sum p_j p_k$ 中, 每項恰好出現一次, 所以 $\sum\sum p_j p_k = (p_1 + p_2 + \cdots)(p_1 + p_2 + \cdots) = 1$. 因此,(4.1)可以作为概率的定义.

設 A 和 B 是原来样本空間 $\mathfrak{S}$ 上任意的两个事件. 用 (A, B) 表示"第一次試驗发生 A 而第二次試驗发生 B". 假設事件 A 含有样本点 $E_{a_1}, E_{a_2}, \cdots$ 而事件 B 含有样本点 $E_{b_1}, E_{b_2}, \cdots$, 則事件 (A, B) 是所有对子 (E_{a_j}, E_{b_k}) 之并. 正如前述,我們有

$$\begin{aligned}\mathbf{P}\{(A,B)\} = \sum\sum p_{a_j} p_{a_k} = (\sum p_{a_j})(\sum p_{b_k}) = \\ = \mathbf{P}\{A\}\mathbf{P}\{B\}.\end{aligned} \tag{4.2}$$

因此,事件 A 与 B 独立. 可見由定义(4.1)出发, 可以推出: 第二次試驗中的任一事件都对第一次試驗中的事件独立. 从概率論的角度来看,这恰好刻画了"相同試驗".

这些考虑也可推广到 r 次試驗的体系,并可导出

定义 1. 令 $\mathfrak{S}$ 是具有样本点 $E_1, E_2, \cdots$ 的一个样本空間, 对应的概率分别为 $p_1, p_2, \cdots$. 所謂对于 $\mathfrak{S}$ 的 r 次独立試驗是指由 r 元組 $(E_{j_1}, E_{j_2}, \cdots, E_{j_r})$ 为样本点的样本空間, 并以如下的方式定义概率

$$\mathbf{P}\{(E_{j_1}, E_{j_2}, \cdots, E_{j_r})\} = p_{j_1} p_{j_2} \cdots p_{j_r}. \tag{4.3}$$

換言之, 新空間的每一点可視为由原空間的点組成(可以重

复）的大小为 r 的一个样本，并且由乘法規律（4.3）来确定新空間上的概率．讀者必須記住：（4.3）并不是定义概率的唯一方法．換句話說，重复試驗并不一定是独立試驗．例如，波利亚罐子模型［例（2.c）］定义了相依（非独立）的重复試驗．等式（4.3）定义了独立試驗，用物理上的話来說，就是在同样条件下的重复試驗．

用导出（4.2）的推論，可以更普遍地用来証明下面定理的正确性．

定理． 設一組事件 $A_1, A_2, \cdots, A_r$，其中 A_j 的发生与否仅由第 j 次試驗单独判断，那么，如果試驗独立，即如果（4.3）成立，則事件 $A_1, \cdots, A_r$ 彼此独立．

若空間 $\mathfrak{S}$ 含有有限多个（設 N 个）样本点，則新空間有 N^r 个样本点 $(E_{j_1}, E_{j_2}, \cdots, E_{j_r})$．若 $\mathfrak{S}$ 的每一点有概率 $\frac{1}{N}$，則由（4.3），点 $(E_{j_1}, \cdots, E_{j_r})$ 的概率是 N^{-r}．新的处理方法想来要比等概的形式上指定还要好，因它可用于非等概的样本空間，也可以用于无穷的样本空間．对于概率的一般理論，这更是一个不可少的工具，因为在一般理論里，那怕我們只做一次单独試驗，我們也把它考慮成一串試驗的第一个．这使得我們总是在考慮着試驗結果的无穷叙列 $(E_{j_1}, E_{j_2}, \cdots)$，并且用一种方法来确定新空間的概率，要定得与（4.3）沒有矛盾．遺憾的是，新空間不是离散样本空間，故这样的討論超出了本卷书所限制的离散样本空間的理論．为了保持本书中理論的初等性，我們所付的代价是：必須按照試驗的次数改变样本空間．

在前面的討論中，我們仅仅考慮同一个試驗的重复进行．但是，不同的試驗的継續也可用同样的方法来处理．倘若我們先扔一个錢币，然后擲一个骰子，則我們自然假設这两次試驗是独立的．这相当于用乘法規律来定义概率．因此有 $\mathbf{P}\{(\text{正面, 么点})\} = \frac{1}{2} \cdot \frac{1}{6}$，等等．在这特殊情形，就等价于定义这所有的 12 个样本点都是等概的．但在一般的情形，必須用（4.3）的方式来定义概率．

定义2. 設 $\mathfrak{S}'$ 与 $\mathfrak{S}''$ 是两个样本空間，分别用 $E_1', E_2', \cdots$ 与 $E_1'', E_2'', \cdots$ 来表示它們的样本点. 并設其相应的概率分别为 $p_1', p_2', \cdots$ 与 $p_1'', p_2'', \cdots$. 則相継的两次試验可以由样本点(E_j', E_k'')所組成的样本空間来描述. 我們說两次相継試驗是独立的，意即用下面的式子来定义概率

$$\mathbf{P}\{(E_j', E_k'')\} = p_j' p_k''. \tag{4.4}$$

[刚才引进的概念在概率論里沒有什么特别意义. 給定两个具有普通的点子 E' 和 E'' 的空間 $\mathfrak{S}'$ 和 $\mathfrak{S}''$，全部的 (E', E'') 所构成的集合称为 $\mathfrak{S}'$ 和 $\mathfrak{S}''$ 的組合积，通常用 $\mathfrak{S}' \times \mathfrak{S}''$ 来表示. 例如，点 (x, y) 所构成的笛卡尔平面就是 x 軸和 y 軸的組合积. (三維空間既可看作 x, y 平面和 z 軸的組合积又可看作三重的組合积.)方程 (4.4) 定义了通常所謂的 $\mathfrak{S}'$ 和 $\mathfrak{S}''$ 中概率的乘积測度. 我們所用的試驗这个詞等价于一个在其中定义了概率的样本空間. 类似地，两次独立試驗的継續可以考虑为对应的样本空間的組合积，其概率由 (4.4) 所定义.

这些概念可以很容易地推广到任意多个空間的乘积中去. 例如，在 (4.3) 中描述了 $\mathfrak{S}$ 自身的 r 維組合积. 这里，初学概率論的人說第一，第二，…次試驗，而一些数学家們則应用术语：第一，第二，… 个坐标空間. (只依賴于第一次試驗的結果的事件也叫作第一个坐标空間中的柱集.)]

所有的数偶 (i, j) (其中 i 和 j 都是介于 1 和 n 之間的整数) 构成了整数集 $1, 2, \cdots, n$ 到自身的乘积. 在无放回抽样中，数偶 (i, i) 是不允許的，因此，大小为 2 的无放回抽样不能直接构成乘积空間. 然而，如下面的例子所証明的，可以用一种不同的方法去表示它，这种方法考虑为相継的独立試驗，而且这种方法可以应用到更复杂的情形中去.

例. (a) 排列. 我們曾考虑由 $a_1, a_2, \cdots, a_n$ 的 $n!$ 个不同排列所組成的样本空間，每个样本点賦以概率 $\frac{1}{n!}$. 这样本空間也可由相継的$(n-1)$次試驗来表达，現叙述如下：先把 a_1 写出，第一次試驗是将 a_2 放在 a_1 的前面或后面，放完 a_2 后，則有三个位置可以放置 a_3，于是，第二次試驗是将 a_3 放到三个位置中的任一个位置，这就决定了 a_1, a_2, a_3 的一个排列. 有四个位置可以放置 a_4. 一般地，当 $a_1, a_2, \cdots, a_k$ 按照某一次序排定之后，我們就进行第 k 次試驗，即把 a_{k+1} 放在 $k+1$ 个位置中的任意一个. 換言之，在相

継的 $(n-1)$ 次試驗中，第 k 次試驗共有 $k+1$ 个可能結果(即样本点)，且每个可能結果有概率 $\frac{1}{(k+1)}$[1]. 由于試驗是独立的，于是，按照原始的定义，n 个元素的任一組排列的概率为 $\frac{1}{2}\cdot\frac{1}{3}\cdots\frac{1}{n}$.

(b) 无放回的抽样. 設总体是 $(a_1, a_2, \cdots, a_n)$. 在无放回的抽样中，每次抽取都拿掉一个元素. 因此，在第 k 次抽取以后，余下的只有 $n-k$ 个元素，而下一次抽取可以由該元素所选择的第 v 号 $(v=1, 2, \cdots, n-k)$ 位置来描述. 如此，无放回地抽取一个大小为 r 的样本变成相継的 r 次試驗，其中，第一次試驗有 n 个可能結果；第二次有 $n-1$ 个；第三次有 $n-2$ 个，… 等等. 在每次試驗中，所有可能的結果都賦以相同的概率. 假定 r 次試驗独立，这就相当于按照随机样本的定义，每个大小为 r 的样本都賦以概率 $1/(n)_r$. [注意，当 $n=100$, $r=3$ 时，样本 (a_{13}, a_{40}, a_{81}) 表示选取的号碼分别为 13, 39, 79. 我們应該說，第三次試驗是从剩下的 $(n-2)$ 个元素所构成的总体中抽取第 79 号元素，而对于原来的号数而言，第三次的試驗結果是与第一次、第二次的結果有关的.] 我們可以看到，重复独立試驗的概念使得我們能把抽样問題的研究看成是相継的个别运算.

*5. 在遺传学中的应用

孟德尔 (G. Mendel, 1822—1884) 創始的遺传学理論对簡单的概率模型的适用性提供了一个富有教益的例解. 下面我們只局限于討論一些最初等的問題. 在描述生物学的背景的时候，我們必須使之簡单化并集中于那些便于数学处理的事实.

遺传性状依賴的特定携带者，称为遺传因子[2]. 除了生殖細胞或者配子以外，人体中的每一个細胞都具有同样的遺传因子結构. 明显的事实是：遺传因子成对出現. 讀者可以把它們設想为一大

---

* 有星号的小节为处理特殊的主题，可以略去，以下相同.

1) 原书錯写为 $1/k$——譯者注.

2) 原文为 gene，遺传学中一般都称为基因——中譯本編者.

堆穿在一段短綫上的珍珠——染色体． 这些染色体也成对出現，而配成对的遺传因子在配成对的染色体上处在相同地位．最簡单的情形是每一对特定的遺传因子中的每一个都只能取两种形式 A 和 a．因此，可以构成三种不同的对子，而且对于这些特定的对子来說，有机体属于这三种遺传型 AA，Aa，aa（在 aA 与 Aa 之間不加区别）．例如，豌豆携带这样一对遺传因子： A 使之开紅花，a 使之开白花．在这种情况下，三种遺传类可以区分为紅花、淡紅色的花和白花． 每一对遺传因子决定一个可遺传的因素，但是机体的可观察到的大多数性質依賴于很多因素． 对于某些特征（例如眼睛的顏色、习慣用左手）来說，某一对特定的遺传因子的影响特別大，在这些情况下，孟德尔定律的效果是容易观察到的．其它一些特征（例如长度），可以設想为一大堆遺传因子的效果的迭加[参看原书第十章例（5.c）]． 这里我們将要研究一对特定的遺传因子的遺传型与継承的問題，对于这特定的一对遺传因子来說，它有三种遺传型 AA，Aa，aa． 一般来說两个遺传因子有 N 种不同的形式 $A_1, \cdots, A_N$，因此就有 $\binom{N+1}{2}$ 个遺传类 A_1A_1，$A_1A_2, \cdots$，A_NA_N．在这种一般的情况下，理論需要作一些修正（参看問題 27）． 下面的計算也可以应用到当 A 具有显性而 a 是隐性的情形中去．这就是說 Aa 这个个体与 AA 具有同样的可观察到的性質，因此，对 a 因子来說，只有純 aa 型才会表現出不同的可观察到的影响．不完全显性的全部演变在自然界中会出現．典型的不完全隐性的性状是蓝眼睛和习慣用左手等等．

生殖細胞或者配子是由分裂过程所产生的而且它只具有一个遺传因子．純 AA 型和 aa 型的机体（或純合子）只产生一种类型的配子，但是 Aa 型的机体（杂合子）产生同等数目的 A 配子与 a 配子． 新的机体由两类上一輩的配子而得来，而且新的机体从双亲的配子得到它的遺传因子．因此每一对都包含一个父系的遺传因子和一个母系的遺传因子，而且每一个遺传因子可以追溯到任何一代特定的祖先，不管是哪一代的远祖．

后代的遺传型依賴于一个随机过程．在每一时刻，每一个上一輩的遺传因子以 1/2 的概率传給后代，而継續的試驗是相互独立的．换句話說，我們可以把第 n 代的遺传型設想为 n 次独立試驗的結果．每一次試驗对应于扔一个錢币．例如，$Aa \times Aa$ 的后代的遺传型配成对时为 AA，Aa，aa，其对应的概率为 $\frac{1}{4}$，$\frac{1}{2}$，$\frac{1}{4}$．一个 $AA \times aa$ 的結合只能产生 Aa 型的后代，等等．

把总体視为全体，我們可以把亲本的配对設想为第二个随机过程的結果．我們将仅仅研究所謂随机交配，它是用下述条件所定义的：如果在第一代子孙中随机选取 r 个，則它們的亲本是所有可能亲本构成的总体中的一个大小为 r 的随机样本．换句話說，每一个后代可以視为随机选取的一对亲本的产物，而且所有这些选取都是相互独立的．随机交配是流行于許多自然界的总体及田間試驗的一些条件的理想模型．然而，如果紅豌豆播种在田地中的某一个角落而白豌豆播种在另外一个角落，則同一种顏色的亲本的交配比随机交配要来得經常些．优先选择性（如象淡顏色优先选择淡顏色）也与随机交配这一条件相违背．完全非随机交配可以用自花受精植物和人工授粉来代表．某些这样的交配系統将要数学地加以分析，但是我們大部分是注意随机交配．

后代的遺传型是四个独立的随机选择的結果．两个亲本的遺传型的选择可以有 3·3 种方式，它們的遺传因子的选择有 2·2 种方式．然而，我們可以把这两个选择結合起来，并且把这个过程描述为这样一个双重的选择：父系的遺传因子和母系的遺传因子的选取是相互独立地、随机地从一个总体里抽取的，这个总体是由父輩的总体中的全部父系所携带的遺传因子和母系所携带的遺传因子的全体所构成．

假定三种遺传型 AA，Aa，aa 在父系和母系中以同样的比例 $u:2v:w$ 而发生．我們将假設 $u+2v+w=1$，而且称 u，$2v$，w 为遺传型的频率．令

$$p = u + v, \qquad q = v + w. \tag{5.1}$$

显然，遗传因子 A 的个数与 a 的个数之比为 $p:q$, $p+q=1$, 我們称 p, q 分别为遗传因子 A 和 a 的频率. 由于在每次选取中，遗传因子 A 被选出的概率为 p, 又因为假定了选取是相互独立的，所以某后代是 AA 的概率为 p^2. 遗传型 Aa 可以用两种方式发生、所以它的概率为 $2pq$. 因此，在随机交配的条件下，一个后代属于 AA, Aa 或 aa 的概率分别为

$$u_1 = p^2, \quad 2v_1 = 2pq, \quad w_1 = q^2. \tag{5.2}$$

例. (a) 所有的亲本都是 Aa(异性結合子)，則 $u=w=0$, $2v=1$, $p=q=\frac{1}{2}$. (b) AA 和 aa 型的双亲具有同样的比数，則 $u=w=\frac{1}{2}$, $v=0$, 而且 $p=q=\frac{1}{2}$. (c) 最后，$u=w=\frac{1}{4}$, $2v=\frac{1}{2}$ 和 $p=q=\frac{1}{2}$. 在这三种情形中，儿輩所对应的 $u_1=\frac{1}{4}$, $2v_1=\frac{1}{2}$, $w_1=\frac{1}{4}$.

为了更好地了解 (5.2) 的意义，让我們固定遺传因子的频率 p 和 q $(p+q=1)$ 来考虑全部遗传型的频率 u, $2v$, w 的系統，其中 $u+v=p$, $v+w=q$. 对第一代子孙来說，一切都导出与 (5.2) 相同的概率. 在它們之中，有下述的特殊分布：

$$u = p^2, \quad 2v = 2pq, \quad w = q^2. \tag{5.3}$$

如果初始的那一代的频率 u, v, w 具有上述特殊关系 (5.3)——如象例 (c) 一样，那么便有：第一代子孙的遗传型的概率为 $u_1=u$, $v_1=v$, $w_1=w$. 因此，我們称形如 (5.3) 的遺传型的分布是平稳的. 对每一个比例 $p:q$ 来說，都对应一个平稳的分布，或者均势.

方程 (5.2) 給出了一个随机选取的第二代个体的遺传型的分布. 在一个大的总体里面，我們期望遗传型的眞实的频率接近于理論分布[1]. 无論父輩的分布 $u:2v:w$ 为何，方程(5.2)定义了一个

---

1) 否則，我們的概率模型就沒有任何实际意义了. 大数定律与中心极限定理給了这个事实以精确的描述，它使我們能够估計随机起伏的影响.

平稳分布，其中遗传因子 A 与 a 分别以频率 $u_1+v_1=u+v=p$ 与 $v_1+w_1=v+w=q$ 出現[参閱(5.1)]. 換句話說，如果观察到的頻率与計算出的概率一致，则第一代子孙将具有平稳的遺传型分布，而且其后代永远具有这一分布. 实际上，偏差会观察得到，但是在大总体中我們可以說：无論父輩这个总体中的結构如何，随机交配所产生的下一代将漸近地具有不变的遺传因子頻率的平稳遺传型分布. 从第二代起，沒有有系統的变化的傾向，第一代子孙就达到稳定的状态. 这首先被哈代[1](G. H. Hardy)所注意到，因此他解决了在孟德尔定律中假定的难点. 特别地，由此推出，在随机交配的条件下，三种遺传型的頻率具有比例 $p^2:2pq:q^2$. 这反过来又可以用来检驗随机交配这一假設.

哈代也指出了，必須对"漸近"一詞加以注意. 甚至对平稳分布来說，我們都可以期望从一代到下一代会发生小的变化. 这就使得我們作如下的設想. 从任何一个父輩的总体开始，用随机交配就建立了下一代中的平稳分布(5.3). 对于一个平稳分布来說，都沒有任何种类的系統改变的傾向. 然而，随机起伏会一代一代地改变遺传因子的頻率，因此遺传結构会慢慢地改动. 不存在一种力量来专門去恢复原始的頻率. 相反，我們的簡化模型导出下述結論(参看原书第十五章例(2.k))：对一个有限总体来說，一个遺传因子最終将要死掉，因此，这个总体最后将属于 AA 或 aa 中的某一个类型. 实际上这并不一定会发生，因为新的遺传因子的产生依賴于变异、选择和其它許多影响. 要研究它，就需要更精細的数学工具(馬尔科夫鏈，扩散理論).

哈代的定理經常被解释为永远可以达到严格的稳定性. 通常犯的錯誤是相信大数定律的作用象是賦与寻求重返原始状态的記性的一种力量，而很多錯誤的結論也是由此假設而导出的.（这里

1) 参閱[34]. 如果先引用一下第九章和第十五章的术語，我們可以把这一情况描述如下. 第 n 代的三个遺传型的頻率是三个随机变量，其期望値由(5.2)所給出，它不依賴于 n. 它們的眞实値从一代到下一代将要发生变化，它构成一个馬尔科夫过程.

所考虑的生物过程是一类重要的馬尔科夫过程，它将在原书第十五章詳細地討論.）注意：哈代的定律不能应用到具有两对遺传因子（例如眼睛的顏色和习慣用左手）9个遺传型 $AABB$, $AABb$, $\cdots$, $aabb$ 的分布中去. 这仍然有平稳分布的傾向，但是在第一代中均势不能达到（参看問題 31）.

*6. 伴性性状

在前一节的导言中曾經說及遺传因子在染色体上. 它們都是成对地出現，而传递却是单个地，因此，在染色体上的全部遺传因子成柱状結合[1]. 所以，我們的遺传因子的遺传方案仍然可以应用到作为单位的染色体上去. 性别决定于两个染色体，阴性是 XX，阳性是 XY. 母亲一定传給一个 X 染色体，而后代的性别决定于父亲传給他（她）的那一个染色体. 因此产生同等数目的阳性和阴性配子. 男孩和女孩出生率之差可以用先天的遺传物的偶然偏差来解释.

我們曾經說过：遺传因子和染色体都是成对出現的. 但是，也有例外的情形，那就是当遺传因子处于 X 染色体上而 Y 染色体上沒有对应的遺传因子的时候. 阴性有两个 X 染色体，因此有两对这样的 X 型的联結的遺传因子；然而在阳性中，X 遺传因子单个出現. 两个伴性的遺传因子会引起色盲和出血素質. 对于它們之中每一个来說，阴性还可以分成三个遺传型 AA、Aa、aa，但是却只有一个遺传因子，而阳性只有两个遺传型 A 和 a. 注意：儿子經常是具有父亲的 Y 染色体，因此，伴性性状不可能由父亲传給儿子. 然而它可以通过父亲传給女儿，再由女儿传給孙子.

現在我們把前一节的分析进行推广. 仍然假定随机交配，并且假定在阴性总体中遺传型 AA, Aa, aa 的頻率分别为 u, $2v$, w. 如前面一样，我們令 $p = u + v$, $q = v + w$. 阳性的两个遺传型 A 与 a 的頻率分别以 p' 和 q' 表示 $(p' + q' = 1)$. 于是 p 和 p' 分

1) 这情形由于染色体的偶然破裂和重新結合而稍复杂些（参閱第二章問題(10.12)）.

别为遺传因子 A 在阴性总体与阳性总体中的頻率. 一个阴性后代的遺传型为 AA, Aa, aa 的概率分别以 u_1, $2v_1$, w_1 表之; 阳性的型 A 和型 a 的类似的概率以 p_1' 和 q_1' 表之. 因为一个阳性后代从其母亲那里得到 X 染色体,所以

$$p_1' = p, \qquad q_1' = q. \tag{6.1}$$

对三个阴性的遺传型来說,如第 5 节一样,我們发現

$$u_1 = pp', \quad 2v_1 = pq' + qp', \quad w_1 = qq'. \tag{6.2}$$

因此

$$\begin{aligned} p_1 &= u_1 + v_1 = \frac{1}{2}(p + p'), \\ q_1 &= v_1 + w_1 = \frac{1}{2}(q + q'). \end{aligned} \tag{6.3}$$

我們可以把这三个公式作如下的解釋. 在阳性的后代中, 屬于母系的遺传因子 A 和 a 分别近似地以頻率 p 和 q 而出現; 而在阴性的后代中, 遺传因子的頻率近似地为 p_1 和 q_1, 或者說在父系与母系中的这些遺传因子的頻率近似为 p_1 与 q_1. 我們发現遺传因子的頻率有趋于相等的趋势. 事实上,从 (6.1) 和 (6.3) 我們得到

$$p_1' - p_1 = \frac{1}{2}(p - p'), \quad q_1' - q_1 = \frac{1}{2}(q - q'). \tag{6.4}$$

这意味着随机交配使下一代的阴性与阳性的遺传因子的頻率之差近似地将要減少一半. 然而,它不能使这个差变为 0, 使之更加減少的趋势是存在的. 与哈代的定律比較, 在一代以后这里沒有平稳的情形. 我們可以追究由一代到一代的变化的系統部分而忽略随机起伏,并驗証理論概率 (6.2) 和 (6.3) 与第一代子孙的对应的实际頻率一致[1]. 对第二代来說,用类似的推导我們可以得到

$$p_2 = \frac{1}{2}(p_1 + p_1') = \frac{3}{4}p + \frac{1}{4}p',$$

1) 如果用第 5 节脚注中所引进的术語, 我們可以把 p_n 与 q_n 解释为第 n 次阴性子孙中遺传因子的期望值. 如果这样解释的話, 关于 p_n 和 q_n 的公式就不再是漸近式而是精确的等式了.

$$q_2 = \frac{1}{2}(q_1 + q_1') = \frac{3}{4}q + \frac{1}{4}q'. \tag{6.5}$$

当然，$p_2' = p_1$, $q_2' = q_1$. 第 n 代阴性子孙中概率 p_n 和 q_n 的一般表达通过几步推导就可以得到. 令

$$\alpha = \frac{1}{3}(2p + p'), \quad \beta = \frac{1}{3}(2q + q'). \tag{6.6}$$

(注意 $\alpha + \beta = 1$.)于是

$$\begin{aligned} p_n &= \frac{p_{n-1} + p_{n-1}'}{2} = \alpha + (-1)^n \frac{p - p'}{3 \cdot 2^n}, \\ q_n &= \frac{q_{n-1} + q_{n-1}'}{2} = \beta + (-1)^n \frac{q - q'}{3 \cdot 2^n}, \end{aligned} \tag{6.7}$$

而 $p_n' = p_{n-1}$, $q_n' = q_{n-1}$. 因此

$$p_n \to \alpha, \quad p_n' \to \alpha, \quad q_n \to \beta, \quad q_n' \to \beta. \tag{6.8}$$

正如(6.2)所給出的一样，在阴性总体中遺传型的頻率为

$$u_n = p_{n-1}p_{n-1}', \quad 2v_n = p_{n-1}q_{n-1}' + q_{n-1}p_{n-1}',$$
$$w_n = q_{n-1}q_{n-1}'. \tag{6.9}$$

因此

$$u_n \to \alpha^2, \quad 2v_n \to 2\alpha\beta, \quad w_n \to \beta^2. \tag{6.10}$$

这些公式說明了，当一代一代传下去的时候，有一个很強的系統趋势，使之最后趋于这样一个情况：遺传型 A 和 a 在阳性中出現的頻率分别为 α 和 β，而阴性的遺传型 AA, Aa, aa 分别具有概率 α^2, $2\alpha\beta$, β^2. 由 (6.7) 看出，其收斂的速度是很快的. 实际上，三四代以后就会达到均势. 确实，上述的变化的随机起伏是很小的，然而后者說明了主导的系統趋势.

我們的主要結論是：在随机交配下，我們可以期望在阳性中的伴性的遺传型 A 和 a；在阴性中的遺传型 AA, Aa, aa 分别漸近地以頻率 α, β, α^2, $2\alpha\beta$, β^2 而发生，其中 $\alpha + \beta = 1$.

应用. 許多伴性遺传因子(如色盲)都是隐性的并且会引起缺陷. 令 a 就是这样的一个遺传因子. 于是所有的阳性的 a 和所有的阴性的 aa 都会显示出缺陷. Aa 型的阴性生物可以把缺陷传給她們的后代，但却不影响她們自己. 因此，我們可以期望，隐

性的伴性缺陷发生在阳性生物中的頻率为 α，而发生在阴性生物中的頻率为 α^2. 如果 100 个男人中有一个色盲的話，那么在 10,000 个女人中才会有一个色盲.

*7. 选　　择

作为选择的影响的一个典型例子，我們将要研究个体 aa 不可能重迭的情形. 当遺传因子 a 是隐性的和致死的时候就会出現这种情形，因此，个体 aa 出生了但是不能生存. 若用人工培育去干涉或阻止个体 aa 之交配时就会出現另外一种情形.

假定在个体 AA 与 Aa 之間进行随机交配，但是不与 aa 型进行交配. 令总体中 AA, Aa, aa 型的遺传型出現的頻率分别为 u, $2v$, w. 于是其亲本对应的頻率为

$$u^* = \frac{u}{1-w}, \quad 2v^* = \frac{2v}{1-w}, \quad w^* = 0. \tag{7.1}$$

我們可以仿照第 5 节一样去进行推导，不过必须以(7.1)中的三个量分别代替 u, $2v$, w. 因此，(5.1) 被

$$p = \frac{u+v}{1-w}, \quad q = \frac{v}{1-w} \tag{7.2}$$

所代替. 第一代子孙中那三个遺传型的概率仍然由 (5.2) 或者說 $u_1 = p^2$, $2v_1 = 2pq$, $w_1 = q^2$ 所給出.

与前面一样，为了研究从一代到一代的系統变化，我們用 u_1, v_1, w_1 分别替代 u, v, w，从而得到第二代的对应的概率 u_2, v_2, w_2, 等等. 一般地，由(7.2)我們得到

$$p_n = \frac{u_n + v_n}{1-w_n}, \quad q_n = \frac{v_n}{1-w_n} \tag{7.3}$$

和

$$u_{n+1} = p_n^2, \quad 2v_{n+1} = 2p_nq_n, \quad w_{n+1} = q_n^2. \tag{7.4}$$

比較(7.3)和(7.4)得到

$$p_{n+1} = \frac{u_{n+1} + v_{n+1}}{1-w_{n+1}} = \frac{p_n}{1-q_n^2} = \frac{1}{1+q_n} \tag{7.5}$$

和

$$q_{n+1} = \frac{v_{n+1}}{1 - w_{n+1}} = \frac{q_n}{1 + q_n}. \tag{7.6}$$

从(7.6)我們可以精确地算出 q_n 来. 事实上

$$\frac{1}{q_{n+1}} = 1 + \frac{1}{q_n}, \tag{7.7}$$

由此相継地得到

$$\frac{1}{q_1} = 1 + \frac{1}{q}, \quad \frac{1}{q_2} = 2 + \frac{1}{q}, \quad \frac{1}{q_3} = 3 + \frac{1}{q}, \cdots,$$

$$\frac{1}{q_n} = n + \frac{1}{q} \tag{7.8}$$

或者

$$q_n = \frac{q}{1 + nq}, \quad w_{n+1} = \left(\frac{q}{1 + qn}\right)^2. \tag{7.9}$$

我們发現不繁殖的遺传型逐漸被淘汰，但是这个过程是很慢的. 对 $q = 0.1$ 来說，到第十代才使遺传因子 a 的頻率减小一半，使 aa 型的頻率近似地减为 $\frac{1}{4}$.（如果 a 是伴性的，則如問題 29 所証明的，其被消除的速度是很快的，而对一般的选择方案来說，請参看問題 30.）

8. 問 題

1. 擲三个骰子. 若已知沒有两面相同，求至少有一个么点的概率.

2. 擲十个骰子. 已知其中至少有一个么点出現，求有两个以上的么点出現的概率 p.

3. 桥牌. 在桥牌游戏中，西家沒有爱司. 問他的合作者（东家）(a) 沒有爱司，(b) 有两个或两个以上的爱司的概率是多少？用直接的推理驗証所得的結果.

4. 桥牌. 設南，北两家共有 10 张王牌（王牌是一种指定花色的牌）. (a) 試求另外 3 张王牌集中在同一家（东家或西家）的概率. (b) 若已知其余 3 张王牌中有一张老开，試求他是"孤家寡人"（意即他只有一张老开，而其余两张王牌在他的合作者手中）的概率.

5. 用例 (2.b) 中的条件概率的概念討論第二章例(7.b)中的鑰匙問題.

6. 在一个螺絲釘制造工厂里，有三个机器 A, B, C，它們的产量分别占

总产量的25%，35%，40%，而它們的产品的废品率分别为5%，4%，2%。今从产品中任意取出一个螺絲釘，而发現它是废品。求这螺絲釘是由机器A或B或C造出的概率。

7. 設100个男人中有5个色盲者，而10,000个女人中有25个色盲者。今从人羣中任取一人，并发現他是色盲。求此人是男性的概率（假設人羣中男女的人数是相同的）。

8. 7个球随机地放入7个盒；各种排列的概率在第二章第5节的表1中都已列出。应用此表，驗証：給定两个盒是空的条件下有一个盒三次地被占有的概率为1/4，精确到5位小数。証明1/4就是正确的答案。

9. 擲一顆骰子，一直到出現么点为止。假定第一次沒擲出么点，問需要擲次数多于3才停止的概率为何？

10（續上）。假定所擲之次数n为偶数。問$n=2$的概率是多少？

11[1]. 設一个家庭恰有n个小孩的概率p_n是αp^n，$n \geq 1$，$p_0 = 1 - \alpha p(1+p+p^2+\cdots)$，并設$n$个小孩的所有可能的性别分布是等概的，証明当$k \geq 1$时，一个家庭恰有$k$个男孩的概率是$2\alpha p^k/(2-p)^{k+1}$。

12（續上）。已知家庭中至少有一男孩，求此家庭至少有两男孩的概率。

13. 骰子A具有四面紅两面白，骰子B具有两面紅四面白。扔一个錢币，如果出現正面，則継續擲骰子A；如果出現反面，則擲骰子B。(a) 証明在任何一次拋擲中，出現紅的概率是1/2。(b) 如果前两次結果是紅的，問第三次擲出紅的概率是多少？(c) 如果前面n次出現紅的，問能擲骰子A的概率是多少？(d) 罐子模型是否和这个游戏等价？

14. 在例子(2.a)中，令x_n为获胜者在第n次試驗获胜的概率，y_n和z_n分别为輸家和等待者在第n次試驗获胜的概率。(a) 証明

$$x_n = \frac{1}{2} + \frac{1}{2} y_{n+1}, \quad y_n = \frac{1}{2} z_{n+1}, \quad z_n = \frac{1}{2} x_{n+1}. \qquad (*)$$

(b) 用直接的簡单推理証明，实質上$x_n = x$，$y_n = y$，$z_n = z$不依賴于n。(c) 証明a胜这一局的概率为5/14（用第一章問題5的推理）。(d) 証明$x_n = \frac{4}{7}$，$y_n = \frac{1}{7}$，$z_n = \frac{2}{7}$是(*)的唯一的有界解。

15. 設事件$A_1, A_2, \cdots, A_n$独立，且$\mathbf{P}\{A_k\} = p_k$，試求沒有一个$A_k (k = 1, 2, \cdots, n)$发生的概率$p$。

---

1) 根据 A. J. Lotka，当$p = 0.7358$时，美国家庭的統計是滿足我們的假設的。参看[35]。

16(續上)。 証明 $p < e^{-\Sigma p_k}$.

17(續上)。 由彭弗雷尼不等式（第四章(5.7)）推出： 在事件 $A_1, \cdots, A_n$ 中有 k 个以上同时发生的概率小于 $(p_1 + \cdots + p_n)^k/k!$.

18.波利亚罐子模型,例(2.c)。 若已知第二次所取出的球是黑的,求第一次所取出的球也是黑的概率.

19.波利亚罐子模型,例(2.c)。 用数学归納法証明：无論是哪一次試驗,黑球出現的概率总是$b/(b+r)$.

20(續上)。 用归納法証明,对任何 $m<n$,第m次与第n次所取出的球是黑,黑或黑,紅的概率分别为

$$\frac{b(b+c)}{(b+r)(b+r+c)}, \quad \frac{br}{(b+r)(b+r+c)},$$

并把此結果推广到多于两次取球的情形.

21.波利亚模型的时間对称性。 設 A, B 表示紅球或是黑球(即 AB 有四种可能的組合方法：紅紅；紅黑；黑紅；黑黑)．証明：已知第 n 次出現 A 后第 m 次出現 B 的概率,是等于已知第 n 次出現 B 后第 m 次出現 A 的概率.

22.在波利亚模型中,設 $p_k(n)$ 是前 n 次取球中出現 k 个黑球的概率,証明下面的遞推关系式

$$p_k(n+1) = p_k(n)\frac{r+(n-k)c}{b+r+nc} + p_{k-1}(n)\frac{b+(k-1)c}{b+r+nc},$$

其中 $p_{-1}(n)$ 解释为零．应用这关系式,即可得到 (2.3) 的一个新証明方法。

23.波利亚分布。 在(2.4)中令

$$\frac{b}{b+r} = p, \quad \frac{r}{b+r} = q, \quad \frac{c}{b+r} = \gamma. \tag{8.1}$$

証明

$$p_{n_1,n} = \frac{\left(-\frac{p}{\gamma}\right)_{n_1}\left(-\frac{q}{\gamma}\right)_{n_2}}{\left(-\frac{1}{\gamma}\right)_n}, \quad n = n_1 + n_2, \tag{8.2}$$

对任何常数(不一定是有理数) $p>0$, $q>0$, $\gamma>0$[1] 都有意义（其中 $p+q=1$)。驗証 $p_{n_1,n}>0$ 和

$$\sum_{\nu=0}^{n}\binom{n}{\nu}p_{\nu,n} = 1^{2)}.$$

因此方程 (8.2) 确定了整数 $0, 1, \cdots, n$ 上的一个概率分布,即波利亚分布。

---

1) 原书为 $\gamma > -1$ 是錯的——校者注.

2) 原书为 $\sum_{\nu=0}^{n} p_{\nu,n} = 1$ 是錯的——校者注.

24.波利亚分布的极限形式. 如果 $n\to\infty$, $p\to 0$, $\gamma\to 0$ 使得 $np\to\lambda$, $n\gamma\to\rho^{-1}$, 则

$$\binom{n}{n_1}p_{n_1,n}\to\binom{\lambda\rho+n_1-1}{n_1}\left(\frac{\rho}{1+\rho}\right)^{\lambda\rho}\left(\frac{1}{1+\rho}\right)^{n_1}[1].$$

驗証这一关系,并証明对于固定的 λ 和 ρ 来說,右边項加起来为 1. (右边是所謂負二項分布;参看第六章第 8 节和第六章問題 37.)

25.用条件概率来解释第二章(11.8).

在生物学中的应用

26.在随机交配的条件下, 总体中属于遺传型 Aa 的个体小于总数的 1/2.

27.把第 5 节中的結果推广到下述情形去: 每一个遺传因子具有 A_1, A_2, $\cdots$, A_k 中的任一形式, 因此有$\binom{k+1}{2}$种遺传型而不是三种遺传型(复等位基因).

28.兄妹交配. 从一个总体中随机地选取二个亲本,假設这个总体中遺传型 AA, Aa, aa 出現的频率分别为 u, $2v$, w. 这个过程一直在他們的后代中重复地进行. 求出第一代第二代第三代子孙的亲本都是属于 AA 的概率[参看第十五章例(2.1)及第十六章例(4.b)].

29.选择. 令 a 为一个隐性的伴性遺传因子,假定在一个选择过程中与阳性的 a 交配是不可能的. 如果在阴性中遺传型 AA, Aa, aa 出現的频率分别为 u, $2v$, w, 試証: 对第一代阴性子孙来說, $u_1=u+v$, $2v_1=v+w$, $w_1=0$, 因此 $p_1=p+\frac{1}{2}q$, $q_1=\frac{1}{2}q$. 这就是說, 在阴性中, 遺传因子 a 減少了一半.

30.第 7 节中的选择問題可以推广到这样一种情况: 假定只有 aa 类的分数 λ $(0<\lambda\leqslant 1)$ 可以除去. 証明

$$p=\frac{u+v}{1-\lambda w},\qquad q=\frac{v+(1-\lambda)w}{1-\lambda w}.$$

更一般地,(7.3)被

$$p_{n+1}=\frac{p_n}{1-\lambda q_n^2},\qquad q_{n+1}=q_n\frac{1-\lambda q_n}{1-\lambda q_n^2}$$

---

1) 原书有錯,它只写 $p_{n_1,n}$,而沒有 $\binom{n}{n_1}$ 这个因子——校者注.

所代替.(这些方程的一般解似乎是不知道的.)

31.同时考虑两对分别具有可能形式 (A, a), (B, b) 的遗传因子. 每一个人传給他的后代以这两点中各一个,并假定这四个可能中的每一种都具有概率 1/4. (当遗传因子在分离的染色体上的时候就是这种情形, 否則它們之間有很强的依賴性.)有 9 种遺传型,我們假定它們在父輩总体中的頻率分别为 U_{AABB}, U_{aaBB}, U_{AAbb}, U_{aabb}, $2U_{AaBB}$, $2U_{Aabb}$, $2U_{AABb}$, $2U_{aaBb}$, $4U_{AaBb}$. 令 $p_{AB} = U_{AABB} + U_{AABb} + U_{AaBB} + U_{AaBb}$, $p_{Ab} = U_{AAbb} + U_{Aabb} + U_{AABb} + U_{AaBb}$, $p_{aB} = U_{aaBB} + U_{aaBb} + U_{AaBB} + U_{AaBb}$, $p_{ab} = U_{aabb} + U_{Aabb} + U_{aaBb} + U_{AaBb}$. 計算第一代子孙的对应的量. 对它来説我們証明: $p_{AB}^{(1)} = p_{AB} - \delta$, $p_{Ab}^{(1)} = p_{Ab} + \delta$, $p_{aB}^{(1)} = p_{aB} + \delta$, $p_{ab}^{(1)} = p_{ab} - \delta$, 其中 $2\delta = p_{AB}p_{ab} - p_{Ab}p_{aB}$. 平稳分布为 $p_{AB} - 2\delta = p_{Ab} + 2\delta$, 等等. (注意: 哈代的定律不能应用;由一代到下一代其結构会发生变化.)

32.假定在某总体中遺传型的頻率为 $u = p^2$, $2v = 2pq$, $w = q^2$. 給定一个具有 Aa 型遺传型的人, 他兄弟也具有同样的遺传型的概率为 $(1+pq)/2$.

注意[1]. 下面这些問題都是建立在家族关系上的, 并且也給出了亲属关系等級概念的一种直观意义. 每一个問題都是前一个問題的継續. 仍旧假定随机交配, 并沿用第 5 节的符号. 我們这里考虑的是一类特殊的馬氏鏈(参看原书第十五章). 矩陣代数可以簡化书写.

33.分别以数字 1, 2, 3 来代替遺传型 AA, Aa, aa, 并令 p_{ik} $(i, k = 1, 2, 3)$ 为已知其父亲(或母亲)的遺传型为 i 的条件下其后代的遺传型为 k 的条件概率. 計算这 9 个概率 p_{ik}, 假定其它的父輩的遺传型为 1, 2, 3 的概率分别为 p^2, $2pq$, q^2.

34.証明在已知其一特定的子女的遺传型为 i 的条件下, 父亲的遺传型为 k 的条件概率也是 p_{ik}.

35.証明: 在已知其孙子(祖父)的遺传型为 i 的条件下,其祖父(孙子)的遺传型为 k 的条件概率为

$$p_{ik}^{(2)} = p_{i1}p_{1k} + p_{i2}p_{2k} + p_{i3}p_{3k}.$$

(矩陣 $(p_{ik}^{(2)})$ 是矩陣 (p_{ik}) 的平方.)

---

1) 第一版在此处有一錯誤,那里用"兄弟"(同父同母)一詞代替了此处的"异父或异母兄弟". 李 (C. C. Li) 和路易斯·沙克斯 (Louis Sacks) 指出了这一錯誤,并給出了正确的公式. 用随机矩陣导出了联合分布及亲属之間的关系. 参看 Biometrika, **40** (1954), 347—360.

36.証明：在已知某人的异父(或异母)弟兄属于遺传型 i 的条件下，此人属于遺传型 k 的条件概率也是 $p_{ik}^{(2)}$.

37.証明：在已知某人的祖父(或孙子)属于遺传型 i 的条件下，此人属于遺传型 k 的条件概率为

$$p_{ik}^{(3)} = p_{i1}^{(2)}p_{1k} + p_{i2}^{(2)}p_{2k} + p_{i3}^{(2)}p_{3k} = p_{i1}p_{1k}^{(2)} + p_{i2}p_{2k}^{(2)} + p_{i3}p_{3k}^{(2)}.$$

(矩陣 $(p_{ik}^{(3)})$ 是矩陣 (p_{ik}) 的三次方. 这种手續給出了亲属关系的等級概念以一种明显的意义.)

38.更一般地,我們定义 $p_{ik}^{(n)}$ 为：在已知某人属于遺传型 i 的条件下,其某一个第 n 代子孙属于遺传型 k 的条件概率. 試用归納法証明：$p_{ik}^{(n)}$ 是下面这一个矩陣的元素

$$\begin{pmatrix} p^2 + \frac{pq}{2^{n-1}} & 2pq + \frac{q(q-p)}{2^{n-1}} & q^2 - \frac{q^2}{2^{n-1}} \\ p^2 + \frac{p(q-p)}{2^n} & 2pq + \frac{1-4pq}{2^n} & q^2 + \frac{q(p-q)}{2^n} \\ p^2 - \frac{p^2}{2^{n-1}} & 2pq + \frac{p(p-q)}{2^{n-1}} & q^2 + \frac{pq}{2^{n-1}} \end{pmatrix}.$$

(这表明：一代一代传下去以后,祖先对子孙的影响逐漸下降,其比例因子为 $1/2$.)

39.再考慮問題 36，不过要把“异父(或异母)兄弟”代之以“亲兄弟”. 証明：这时对应的矩陣为

$$\begin{pmatrix} \frac{1}{4}(1+p)^2 & \frac{1}{2}q(1+pq) & \frac{1}{4}q^2 \\ \frac{1}{4}p(1+p) & \frac{1}{2}(1+pq) & \frac{1}{4}q(1+q) \\ \frac{1}{4}p^2 & \frac{1}{2}p(1+q) & \frac{1}{4}(1+q)^2 \end{pmatrix}.$$

40.証明叔侄之間的亲戚关系的等級与祖父和孙子之間的亲戚关系的等級是一样的.

第六章　二項分布与普阿松分布

1. 伯努利試驗序列[1]

在重复的独立試驗中，如果每次試驗仅有两个可能結果，而且其相应的概率在每次試驗中都是相同的，則称这一串重复的独立試驗是伯努利試驗序列．我們經常以 p, q 表示两个相应的概率，并将具有概率 p 的可能結果称为“成功”，記以 S，而另一个可能結果則称为“失敗”，記以 F．显然，p 与 q 应該是非負的，并有

$$p + q = 1. \tag{1.1}$$

每次試驗所对应的样本空間是由两个样本点 S 与 F 組成的，而 n 次伯努利試驗的样本空間則含有 2^n 个样本点，它就是由 n 个符号 S 与 F 所組成的排列．每个排列表示联合試驗中的一个可能結果．因为試驗是独立的，所以它們的概率是等于各个概率的乘积．換言之，每一給定排列的概率可由下面的方法得出，在排列中用 p、q 分別代替 S, F 后，所得的一个乘积便是所要的概率，因此

$$\mathbf{P}\{SSFSF\ldots FFS\} = ppqpq\cdots qqp.$$

例． 伯努利試驗序列的一个最熟悉的例子是連續扔一个均匀的錢币，此时 $p = q = \frac{1}{2}$．如果錢币是不均匀的，且我們假設連續扔錢币的試驗是独立的，則我們也得到了伯努利試驗序列的模型，不过此时成功的概率 p 可以是任何数值[2]．每次都从一个包含有 r 个紅球 b 个黑球的罐子中重复地随机取球，这种取球代表一串具有 $p = r/(r+b)$ 的伯努利試驗．有时，一个試驗会有好

---

1) James Bernoulli (1654—1705)，他的主要工作“Ars conjectandi”于 1713 年出版．

2) 这里应該說是 0 与 1 之間的任意数值——譯者注．

几个可能的結果，但我們并不注意各个結果的区別，而是把这些結果簡单地区分为两組 A 或非 A. 例如，擲一个均匀的骰子，则么点出現 (S) 与非么点出現 (F) 的区别便导至一个伯努利試驗，此时 $p=\frac{1}{6}$; 偶数点出現与奇数点出現的区别也导至一个成功概率 $p=\frac{1}{2}$ 的伯努利試驗. 若骰子是不均匀的，則重复擲骰子也組成一串伯努利試驗，但是成功的概率 p 可有各种不同的数值. 扑克中的最大同花或擲两个骰子都出么点都可以用成功来表示，而所有其他的可能結果的出現用失敗表示，則我們有成功概率分别为 $p=1/649,740$ 与 $p=\frac{1}{36}$ 的伯努利試驗. 在統計应用中常常把問題化为这种类型来处理. 例如，在大量制造垫螺釘的小鉄片时，鉄片的厚度可有各种不同的数值，但是，在产品检查中，我們往往把鉄片的厚度分为两类，即如果厚度在預先規定的范围內，則称为合格 (S); 反之則称为不合格 (F).

伯努利試驗序列是一个理論的模型，并且，用它来描述某一指定的实际試驗是否合适仅能由經驗来判断. 关于連續扔一个錢币是伯努利模型的知識是由实驗的証据推导出来的. 街上的行人和哲学家馬伯[1] (K. Marbe) 都相信在連續出現 17 次的"正面"后，"反面"的出現将是更概然的. 这个主张并不牵涉到錢币完善不完善的問題，而是把記忆力付与了自然界，也就是（用我們的語言来說）否認了逐次試驗的独立性. 馬伯的理論不能用邏輯来駁倒，但是，由于它缺乏經驗的支持，人們还是拒絕接受它.

在抽样的实践、工业的質量控制等等中，伯努利試驗序列的模型提供了一个理想的标准，虽然这标准从来不能完全达到. 在上面所講的垫螺釘的鉄片制造的例中，有許多理由說明为什么产品不能形成伯努利模型. 例如，由于机器的必然起的变化使得概率不能保持为常数. 机器的运轉有慣性的傾向，因此，同类偏差的长

1) 参看[36]，关于馬伯的理論，有許多批評的文章.

連貫的出現，是要比各次动作純为独立的情况下更为概然的．但是，从質量控制的观点来看，我們愿意生产过程符合一个伯努利模型，而一个重要的发現是，在一定范围之內，生产过程是可以达到这个要求的．这样一来，連續控制的目的，就是要及早地发現显著地与理想模型的偏离，并利用它作为就要发生毛病的警报．

2. 二 項 分 布

在 n 次伯努利試驗中，我們往往只关心成功的总次数而不計較成功的排列次序．由于成功的总次数可能是 $0, 1, \cdots, n$，因此，我們的第一个問題是求出它們相应的概率．但在"n 次試驗里有 k 次成功，$n-k$ 次失敗"的出現可以有許多的方式，其方式的个数就等于 k 个文字 S 在 n 个位置上的所有可能分配法，換言之，这一事件共含有 $\binom{n}{k}$ 个样本点，并且，由定义，每点的概率都是 $p^k q^{n-k}$．于是，我們証明了下面的定理．

定理． 在成功概率为 p，失敗概率为 $q=1-p$ 的 n 次伯努利試驗中，有 k 次成功，$n-k$ 次失敗 $(0 \leqslant k \leqslant n)$ 的概率 $b(k; n, p)$ 是

$$b(k; n, p) = \binom{n}{k} p^k q^{n-k}. \tag{2.1}$$

特別，未成功的概率是 q^n，而至少有一次成功的概率是 $1-q^n$．

我們考虑 p 为常数，而且令 $\mathbf{S}_n$ 为 n 次試驗的成功次数；則 $b(k; n, p) = \mathbf{P}\{\mathbf{S}_n = k\}$．用一般的术語来說，$\mathbf{S}_n$ 是随机变量，函数 (2.1) 是这个随机变量的"分布"；我們将称这个分布为二項分布．因为 (2.1) 是 $(p+q)^n$ 的二項展开式中的第 k 項[1]，所以我們有"二項"两字的名称．这个解释还指出来 $b(0;n,p)+b(1;n,p)+\cdots+b(n;n,p)=(p+q)^n=1$．这也正是概率的概念所需要的．二項分布有表可查[2]．

---

1) 此处有錯，应为第 $k+1$ 項——譯者注．

2) 对 $n \leqslant 50$，参看 [37]；对 $50 \leqslant n \leqslant 100$ 参看 [38]；对更大的区域，則参看[39]．

例. (a) 威尔頓(Weldon)擲骰子記录表. 設所进行的試驗是擲 12 个骰子,并設出現点数 5 与 6 是"成功". 如果骰子是完善的,则成功的概率 $p=\frac{1}{3}$. 于是,成功次数的概率将由二項分布 $b\left(k;12,\frac{1}{3}\right)$ 来表达. 表 1 給出了这些概率以及在 26,306 次的实际实驗中所得到的相应的观察平均頻率. 这两者的相合性看起来似乎是不錯的. 但是,对于如此大量的数据来說,符合程度其实是很不够的. 統計学家常常用 χ^2 准則来判断拟合的程度. 根据这个准則,对于均匀的骰子而言,象威尔頓观察到的那样大的偏差只有万分之一的发生概率. 因此,我們有理由假設骰子是有偏的. 当成功的概率改为 $p=0.3377$ 时,最符合于我們的观察結果[40].

表 1. 威尔頓擲骰子記录表

k	$b(k;12,1/3)$	观察的頻率	$b(k;12,0.3377)$
0	0.007 707	0.007 033	0.007 123
1	.046 244	.043 678	.043 584
2	.127 171	.124 116	.122 225
3	.211 952	.208 127	.207 736
4	.238 446	.232 418	.238 324
5	.190 757	.197 445	.194 429
6	.111 275	.116 589	.115 660
7	.047 639	.050 597	.050 549
8	.014 903	.015 320	.016 109
9	.003 312	.003 991	.003 650
10	.000 497	.000 532	.000 558
11	.000 045	.000 152	.000 052
12	.000 002	.000 000	.000 002

(b) 在第四章第 4 节当我們考虑牌的猜測問題时,曾經遇到过二項分布,那里的表 3 中 b_m 諸列列出了 $n=3,4,5,6,10$ 和 $p=\frac{1}{n}$ 时分布中的各項. 在第二章(4.c)的占位問題中,我們找出了公式(4.5),它是二項分布的另一个特殊情形.

(c) 如果成功的概率是 0.01,問需要有多少次試驗才能使得

至少出現一次成功的概率是 1/2 或更大？ 此时，我們是求最小的 n，使得 $1-(0.99)^n \geq \frac{1}{2}$，或是 $-n\log(0.99) \geq \log 2$，因此 $n \geq 70$.

(d) 能量供应問題. 假定有 $n=10$ 个工人間歇性的使用电力，我們感兴趣的是估計所需要的总負荷. 作为一种粗糙的近似，我們設想在任何一个給定时刻每一个工人以同样的概率 p 需要一个单位电力. 如果他們是独立地进行工作，則恰有 k 个工人同时需要电能的概率是 $b(k; n, p)$. 如果一个工人在一小时內平均有 12 分鈡需要电能，則我們令 $p=\frac{1}{5}$. 于是在同时有七个或者七个以上的工人需要电能的概率为 $b(7; 10, 0.2)+\cdots+b(10; 10, 0.2)=0.0008643584$. 換句話說，如果最多只能供应 6 个单位电力，則超过負荷的概率为 $0.00086\cdots$，卽是 1157 分鈡內約有一分鈡，亦卽約 20 个工作时中可能有一分鈡超过負荷. 八个或八个以上的工人同时需要电能的概率仅仅为 0.0000779264，卽比上述概率的 1/11 还要小.

(e) 血清或防疫检驗[41]. 假定某种疾病在牲畜中传染的正規比例是百分之二十五. 为了检驗一种新发現的血清，我們把 n 个健康的牲畜都注射这种血清. 我們如何用数字来表示这一試驗的結果呢？ 如果这种血清完全不起作用，則这 n 个被注射的牲畜中恰有 k 个不感染这种疾病的概率为 $b(k; n, 0.75)$. 当 $k=n=10$ 时，这个概率大約为 0.056，而当 $k=n=12$ 时，这个概率只有 0.032. 因此，如果 10 个或者 12 个牲畜中沒有一个感染这种疾病，那么我們就可以用它指明这种血清是有效的，虽然这还不能作为最后的論断. 注意：如果不注射血清，17 个牲畜中至多只有一个感染这种疾病的概率大約为 0.0501. 因此，17 个被注射的牲畜中只有一个感染病比 10 个被注射的牲畜中沒有一个感染病更能說明这种血清有效. 当 $n=23$ 时，至多有两个牲畜感染疾病的概率大約为 0.0492，因此，23 个牲畜中只有两个感染疾病又比 17 个中只有一个感染疾病，10 个中沒有一个感染病更能說明血清之有效性.

(f) 另一种統計检驗. 假定 n 个人服用某种药后量他們的血压，不服用时也量他們的血压. 因此，得到两組測量結果 $x_1, \cdots, x_n$ 和 $x'_1, \cdots, x'_n$. 如果 $x_i < x'_i$，則我們説第 i 次試驗的結果是成功，如果 $x_i > x'_i$，則説第 i 次試驗的結果是失敗.（为了簡单起見，我們假定沒有两个測量的結果相等.）如果这种药沒有效果，則我們的观察对应着具有概率 $p = \frac{1}{2}$ 的 n 次伯努利試驗，成功的次数很多，就可以作为这种药是有效的一个証据.

3. 中心項及尾項

从 (2.1) 我們看到

$$\frac{b(k; n, p)}{b(k-1; n, p)} = \frac{(n-k+1)p}{kq} = 1 + \frac{(n+1)p-k}{kq}. \quad (3.1)$$

因此，当 $k < (n+1)p$ 时，$b(k; n, p)$ 比 $b(k-1; n, p)$ 大；当 $k > (n+1)p$ 时，$b(k; n, p)$ 比 $b(k-1; n, p)$ 小. 如果 $(n+1)p = m$ 是整数，則 $b(m; n, p) = b(m-1; n, p)$. 恰巧存在一个整数 m 满足

$$(n+1)p - 1 < m \leqslant (n+1)p, \quad (3.2)$$

因此我們有

定理 1. 当 k 由 0 变到 n 时，$b(k; n, p)$ 最初单調增加，而后单調下降，而且在 $k = m$ 达到它的最大値，但当 $m = (n+1)p$ 时，$b(m-1; n, p) = b(m; n, p)$.

我們称 $b(m; n, p)$ 为中心項. m 通常称为"成功的最大概然次数"，但是，必須了解，当 n 很大时，所有的 $b(k; n, p)$ 都很小. 把一个均匀的錢币扔 100 次，正面出現的最大概然次数是 50，但是正面出現 50 次的概率比 0.08 还小. 在下一章里，我們将要証明 $b(m; n, p)$ 近似地等于 $(2\pi npq)^{-\frac{1}{2}}$.

显然，(3.1) 里的比値当 k 增加时单調下降，因此，当 $k \geqslant r+1$ 时，我們有

$$\frac{b(k; n, p)}{b(k-1; n, p)} < \frac{(n-r)p}{(r+1)q}. \quad (3.3)$$

令 $k = r + 1, \cdots, r + v$, 再把这 v 个不等式相乘, 卽得

$$\frac{b(r+v;n,p)}{b(r;n,p)} < \left\{\frac{(n-r)p}{(r+1)q}\right\}^{v}. \tag{3.4}$$

当 $r \geqslant np$ 时, 括号內的分数小于 1, 因此把它对 v 求和, 卽得一个公比为 $(n-r)p/(r+1)q$ 的有限几何級数. 因此, 对 $r \geqslant np$, 我們断言

$$\sum_{v=0}^{n-r} b(r+v;n,p) < b(r;n,p)\frac{(r+1)q}{r+1-(n+1)p}. \tag{3.5}$$

左边是二項分布的右端的"尾項", 卽是至少成功 r 次的概率. 把同样的計算应用到左端的"尾項"上去, 可以証明, 对于 $s \leqslant np$ 有

$$\sum_{\rho=0}^{s} b(\rho;n,p) < b(s;n,p)\frac{(n-s+1)p}{(n+1)p-s}. \tag{3.6}$$

因此, 我們証明了

定理 2. 如果 $r \geqslant np$, 至少成功 r 次的概率满足不等式 (3.5); 如果 $s \leqslant np$, 最多成功 r 次的概率满足不等式 (3.6).

[另一种証明, 参看問題 39 (a).]

4. 大 数 定 律

我們曾經几次地提到过概率的直观概念是基于以下的假定上的. 如果在 n 次相同的試驗中, A 发生 v 次, 而且令 n 很大, 則 v/n 将接近于 A 的概率 p. 显然, 一种形式上的数学理論从来不可能与現实的生活直接联系起来, 但是它至少应該对我們所要解释的現象提供理論的描象. 因此, 我們要求把刚才說的那句不够清楚的話加以精确化, 而叙述为定理的形式. 为此, 我們理解"相同的試驗"为成功概率是 p 的"伯努利試驗". 如果 $\mathbf{S}_n$ 是 n 次試驗中成功的次数, 則 $\mathbf{S}_n/n$ 是成功的平均次数而且接近于 p. 現在, 我們可以很容易給它一个精确的意义. 例如, 考虑 $\mathbf{S}_n/n$ 超过 $p+\varepsilon$ (其中 $\varepsilon > 0$ 是一个任意小但是固定的常数) 的概率. 这个概率就是 $\mathbf{P}\{\mathbf{S}_n > n(p+\varepsilon)\}$, 它等于(3.5)的左边(这时 r 是超过 $n(p+\varepsilon)$ 的最小整数). 因此, (3.5)蘊含

$$\mathbf{P}\{\mathbf{S}_n > n(p+\varepsilon)\} < b(r;n,p)\frac{n(p+\varepsilon)+q}{n\varepsilon+q}. \quad (4.1)$$

当 n 增加时，右边的分数保持有界，此外因为对每一个滿足 $(n+1)p \leqslant k < r$ 的 k，都有 $b(r;n,p) < b(k;n,p)$，这样的 $b(k;n,p)$ 大約有 $n\varepsilon$ 項，所以 $b(r;n,p) \to 0$. 这就推出，当 n 增加时，$\mathbf{P}\{\mathbf{S}_n > n(p+\varepsilon)\} \to 0$. 应用公式(3.6)，用同样的方法可得 $\mathbf{P}\{\mathbf{S}_n < n(p-\varepsilon)\} \to 0$，因此，我們有

$$\mathbf{P}\left\{\left|\frac{\mathbf{S}_n}{n} - p\right| < \varepsilon\right\} \to 1. \quad (4.2)$$

用文字表示，当 n 增加时，成功的平均数与 p 的偏差超过預先指定的 ε 的概率趋于 0. 这是大数定律的一种形式，并且它是把概率作为相对頻率的一个度量的直观概念的基础. 对于实际应用必需以(4.2)左边的概率的更精确估計来充实；这样一个估計将在二項分布的正态逼近中給出[参看第七章例(3.g)]. 实际上公式(4.2)是第七章問題 18 的一个簡单的推論.

結論(4.2)是古典的大数定律. 它的用处是非常有限的，我們将要用更精确更有用的強大数定律来代替它（参看第八章第 4 节）.

注意. 常常把許多不能从大数定律推出的事件也訴諸于大数定律. 如果甲和乙扔一个錢币 10,000 次，常常这样希望：甲領先的次数大約为一半. 但是，这是不对的. 反正弦律(第三章第 5 节)指出：領先次数各半的可能很小. 甲領先的次数小于 20 的概率比領先次数落在 4990 和 5010 之間的概率大得多. 不存在領先期間相等的任何傾向. 大数定律只是說：在大量的不同的扔錢币的游戏中，在任何一个給定的时刻正面領先的頻率近似地为 1/2，至于在一个固定的游戏中，領先的起伏問題，大数定律什么也沒有說.

5. 普阿松逼近[1]

在很多应用問題中，我們常常遇到这样的伯努利試驗，其中，

---

1) Siméon D. Poisson (1781—1840). 参看他著的书[42].

相对地說，n 大，p 小，而乘积

$$\lambda = np \tag{5.1}$$

是大小适中的．在这种情形下，我們采用 $b(k; n, p)$ 的一个近似公式是很方便的，这近似式是由普阿松得出的．下面，我們将推导这公式．由于 $b(0; n, p) = (1-p)^n$，現将 (5.1) 代入其中，則得

$$b(0; n, p) = \left(1 - \frac{\lambda}{n}\right)^n. \tag{5.2}$$

取对数并且利用泰劳 (Taylor) 展开式［第二章公式 (8.10)］，我們有

$$\log b(0; n, p) = n \log\left(1 - \frac{\lambda}{n}\right) = -\lambda - \frac{\lambda^2}{2n} - \cdots, \tag{5.3}$$

因此，对于充分大的 n，有

$$b(0; n, p) \approx e^{-\lambda}, \tag{5.4}$$

其中符号 $\approx$ 表示漸近相等（在現在这种情形下，它們相差阶为 n^{-1} 的无穷小量）．此外，从 (3.1) 得知：对任一个固定的 k 和充分大的 n 都有

$$\frac{b(k; n, p)}{b(k-1; n, p)} = \frac{\lambda - (k-1)p}{kq} \approx \frac{\lambda}{k}. \tag{5.5}$$

当 $k = 1$，从(5.5)和(5.4)我們得出 $b(1; n, p) \approx \lambda e^{-\lambda}$．当 $k = 2$，我們得到 $b(2; n, p) \approx \lambda^2 e^{-\lambda}/2$．一般地，由归納法我們得到

$$b(k; n, p) \approx \frac{\lambda^k}{k!} e^{-\lambda}. \tag{5.6}$$

这就是著名的二項分布的普阿松逼近．（在問題 30—34 中，給出了它的誤差估計，并且証明了当 $n \to \infty$ 和 $p \to 0$ 使得 $\lambda = np$ 保持有界时，(5.6) 的逼近是一致的．）为了方便起見，把 (5.6) 的右端引进一个符号，我們令

$$p(k; \lambda) = e^{-\lambda}\frac{\lambda^k}{k!}. \tag{5.7}$$

采用这个符号，我們可以說：当 n 充分大时，$p(k; \lambda)$ 是 $b\left(k; n, \frac{\lambda}{n}\right)$ 的近似表达式．

例. (a) 第四章表 3 中最后一列的記录 p_m 給出了 $p(m;1)$ 的数值. 在前面几列中, b_m 却給出了 $b\left(m;N,\frac{1}{N}\right)$的数值. 这个表使我們能够比較普阿松分布 $p(m;1)$ 与具有 $p=\frac{1}{n}(n=3,4,5,6,10)$的二項分布. 比較的結果告訴我們,卽使当 n 很小时, 它們的相合性也惊人的好.

(b) 表 2 比較了普阿松分布 $p(k;1)$ 与 $n=100$, $p=\frac{1}{100}$ 的二項分布. 它表明在很多应用里近似程度是令人滿意的. 現在考虑 100 对随机数字中 (7, 7) 出現的次数 k, k 的分布就是二項分布 $b\left(k;100,\frac{1}{100}\right)$. 表 2 [43]中最后一列給出了每組为 100 对随机数字的 100 組中 (7, 7) 出現的次数. 以 100 除最后一行的所有記录, 我們就得到 (7, 7) 出現次数的頻率, 而这些頻率与理論概率的一致性是合理的. [由 χ^2 标准判断, 在 100 次同样的情形中, 大約 75 次会由于随机摆动而产生观察頻率与理論概率的更大的偏离.]

表 2. 普阿松近似的一个例子

k	$b\left(k;100,\frac{1}{100}\right)$	$p(k;1)$	N_k
0	0.366 032	0.367 879	41
1	.369 730	.367 879	34
2	.184 865	.183 940	16
3	.060 999	.061 313	8
4	.014 942	.015 328	0
5	.002 898	.003 066	1
6	.000 463	.000 511	0
7	.000 063	.000 073	0
8	.000 007	.000 009	0
9	.000 001	.000 001	0

头三列表明了二項分布的普阿松近似. 最后一列是每組为 100 对随机数字的100組中 (7, 7) 出現次数的記录.

(c) 生日． 在 500 人所組成的人羣中，恰有 k 个人的生日是元旦的概率 p_k 是多少？如果 500 人是任意选取的，則我們可应用成功概率为 $p=\frac{1}{365}$ 的 500 个伯努利試驗的模型．于是有 $p_0=(364/365)^{500}=0.2537\cdots$．对于普阿松近似式，我們設 $\lambda=500/365=1.3699\cdots$．于是 $p(0,\lambda)=0.2541$，它所产生的誤差仅在第四位小数上．当 $k=1,2,\cdots$，p_k 的正确数值由二項公式算出如下：$p_1=0.3484\cdots$，$p_2=0.2388\cdots$，$p_3=0.1089\cdots$，$p_4=0.0372\cdots$，$p_5=0.0101\cdots$，$p_6=0.0023\cdots$，其相应的普阿松近似值是 $p(1;\lambda)=0.3481\cdots$，$p(2;\lambda)=0.2385\cdots$，$p(3;\lambda)=0.1089\cdots$，$p(4;\lambda)=0.0373\cdots$，$p(5;\lambda)=0.0102\cdots$，$p(6;\lambda)=0.0023\cdots$．所有的誤差都在第四位小数上．

(d) 废品． 設螺絲釘的生产是在統計的質量控制下进行，于是容許我們应用伯努利試驗的模型．若一个螺絲釘是废品的概率为 $p=0.015$，則装有 100 个螺絲釘的盒中沒有废品的概率是 $(0.985)^{100}=0.22061$．其对应的普阿松近似值是 $e^{-1.5}=0.22313\cdots$．在大多数的实际問題中，这样的近似程度已是相当好的．現在，我們問：一盒中应有多少个螺絲釘才能使得其中含有 100 个合格釘的概率大于或等于 0.8？設 $100+x$ 是所需要的螺絲釘数，則 x 必是很小的整数，应用普阿松近似式于 $n=100+x$，$\lambda=np$，由于 np 是近似地为 $100p=1.5$，故我們所求的 x 是满足下面不等式的最小整数 x：

$$e^{-1.5}\left\{1+\frac{1.5}{1}+\cdots+\frac{(1.5)^x}{x!}\right\}\geq 0.8. \tag{5.8}$$

查表[1] 知，当 $x=1$ 时，(5.8) 的左边是近似于 0.56；而 $x=2$ 时，則为 0.809．因此，从普阿松近似式可得出如下的結論，即所需要的螺絲釘是 102 个．因 0.809 是很靠近所給定的要求 0.8，故 102 个还是不太保险的，而 103 个是較保险一些．实际上，在一盒

1) 参看[44]．(这是給出数值 $p(k;\lambda)$ 与 $p(k;\lambda)+p(k+1;\lambda)+\cdots$ 的一些表格，其中 k 从 0 变到 100．)

102个螺絲釘中,含有100个合格釘的概率是

$$(0.985)^{102}+\binom{102}{1}(0.985)^{101}(0.015)+$$
$$+\binom{102}{2}(0.985)^{100}(0.015)^2=0.8022\cdots.$$

(e) 百岁的人. 某一指定的初生者活到100岁的概率是很小的,而同时,在一个大的团体中,一年的出生数却是大的. 由于战爭,传染病等等,不同人的寿命不是随机独立的. 不过,作为首次的近似,我們可将 n 个出生者譬喻为 n 个伯努利試驗,其中活到100岁以后才死的就算"成功". 在一个稳定的团体、也即团体的大小与死亡率訒为是不大改变的,我們有理由期望,恰有 k 个百岁老人死掉的那些年头的頻率是近似于 $p(k;\lambda)$,其中 λ 依賴于团体的大小及健康情况. 瑞士的記录証实了这結論[45].

(f) 印錯字,葡萄干,等等. 設在一本书的印刷中,如果一个字印錯的概率是固定的,又如果印刷的条件保持不变,则我們有与字的个数一样多的伯努利試驗,并且恰有 k 个错字的頁数的頻率将近似地等于 $p(k;\lambda)$,其中 λ 依賴于印刷者的技术水平. 印刷者的偶然疲倦或困难的章节等等都将增加印錯的机会,并且还可能产生成羣的錯字. 因此,普阿松公式可用来发現对于均匀性或統計控制的过分偏离. 类似的推理可用于許多其他的情形. 例如,若将許多葡萄干混合到一堆生面团中,則我們可期望恰有 k 粒葡萄干的面包的个数的頻率是近似地等于 $p(k;\lambda)$,而 λ 是生面团中葡萄干的密度.

6. 普阿松分布

在前一节中,我們仅仅用普阿松表达式(5.7)作为 n 大 p 小的二項分布的一个合适的近似式. 关于第四章的相合問題与占位問題,我們曾經研究过各种不同的概率分布,而这些分布的极限形式也是普阿松表达式 $p(k;\lambda)$. 这正是下述的值得注意的事实的特殊情形: 有几个具有高度普遍性的分布在极其多种多样的問題中

出现．在概率論中常遇到的三个主要分布是二项分布，正态分布（于下一章介紹）与普阿松分布

$$p(k;\lambda)=e^{-\lambda}\frac{\lambda^k}{k!}, \tag{6.1}$$

后者我們将加以討論．

我們注意，如果将(6.1)的右边对 $k=0,1,2,\cdots$ 求和，則得到 $e^{-\lambda}$ 与 e^{λ} 的泰劳級数的乘积．因此，对任何固定的 λ，所有的 $p(k;\lambda)$ 的和是 1．于是，我們可以設想一个理想的試驗，使得恰有 k 个成功的概率是 $p(k;\lambda)$．我們将說明：为什么許多物理試驗与統計观察都导至普阿松分布（6.1）的如上的解释．下一节的例子闡明了（6.1)的广泛的、种种重要的应用．只能在随机过程的理論中才明显地看出普阿松分布的实质（参看原书 17 章，它給出普阿松分布的一个新的描述)．

考虑一个随时間而发生的随机事件序列，例如放射性分裂；来到电話总机的呼喚等等．每个事件都可由时間軸上的一点来代表，而我們所关心的是这些点的随机分布．这分布有各种不同的类型．但它們的研究是属于連續概率的范围，我們預备在第二卷再談．这里，我們将滿足于說明以下的事实：在最簡单的物理假設下可以得出：$p(k;\lambda)$ 就是“在一个指定长度的固定区間內有 k 个点(事件)”的概率．本节中，我們所用的方法是粗糙的．在第十七章，我們将以更适当的方法来处理这个問題．

我們需要把以下两个物理假定用数学語言表达出来：1）对时間而言，試驗是在固定不变的情况下进行的；2）在不相交的几个时間区間中，事件的发生次数是随机独立的．在連續型的概率理論中，我們可以直接地表达这些陈述，但由于限制在离散型的范围以內，我們只得先用一个近似的有限模型，然后再过渡到极限．

我們想象把单位时間区間分成为数很大的 n 等分，每个子区間的长度是 $\frac{1}{n}$．于是，一个指定的子区間或是空的或是至少含有一个随机点（事件)．我們分别称这两个可能的情形为失敗与成

功．由于 n 个子区間的长度是同样的，因此，在每个子区間中，成功的概率 p_n 也应該相同．还有，不相交区間的独立性使我們能够把这种現象考虑为 n 个伯努利試驗，其中恰有 k 个成功的概率是 $b(k;n、p_n)$．因为一个子区間可以含有若干个随机点，所以成功的次数与随机点的个数就不一定相同．但是，我們可以很自然地引进一个附加的假設，即在一个很短的时間区間內出現两个或多个随机点的概率是可以忽略的[1]．在这假設下，单位时間区間中恰有 k 个随机点的概率可由 $n\to\infty$ 时 $b(k;n,p_n)$ 的极限来表出．現在，我們把每个子区間再分成两等分，則有 $p_n=2p_{2n}-p_{2n}^2$．（因为在长度为 $\frac{1}{n}$ 的区間发生成功意味着在左半边，或右半边，或左、右两半边发生成功．）于是有 $p_n<2p_{2n}$．这使我們想到，np_n 是单調增加的（我們可以严格証明这点）．若 $np_n\to\lambda$，則 $b(k;n,p_n)\sim b(k;n,\lambda/n)\to p(k;\lambda)$．于是，(6.1) 可作为单位时間区間內恰有 k 个随机点发生的概率．倘若我們假定 $np_n\to\infty$，就得不到有意义的結果．因为这隐含着在无論多短时間內可以有无穷多个随机点．

如果以长度为 t 的区間来代替单位区間，并且把单位区間分成許多长度为 $\frac{1}{n}$ 的子区間，那么，我們仍有一串成功概率是 p_n 的伯努利試驗，不过試驗的次数不是 n，而是最靠近 nt 的那个整数．过渡到极限后还是一样，不过是把 λ 换成 λt 而已．因此，我們考虑

$$p(k;\lambda t)=e^{-\lambda t}\frac{(\lambda t)^k}{k!} \tag{6.2}$$

作为长度为 t 的固定区間內恰有 k 点发生的概率．特别，在长度为 t 的区間內沒有一点发生的概率是

---

1) 用这个假設时須要注意：那随机事件不要是成对发生的．举例来說，“車禍”的事件不是成对发生的．我們的假設就能适用：在很短的时間內发生两次的概率比发生一次的概率小得多．但是“車被撞毀”的事件就是成对发生的，因为車总是被別的車撞毀的，而別的車也不免撞毀了，于是我們的假設就不适用．

$$p(0;\lambda t)=e^{-\lambda t},\tag{6.3}$$
从而,有一点或多点发生的概率是 $1-e^{\lambda t}$.

参数 λ 是一个物理常数，它是由 t 軸上随机点的密度来决定的. λ 愈大,沒有点出現的概率 (6.3) 愈小. 設一个物理試驗大量的重复进行例如說 N 次，并且每次我們都記录下在长度为 t 的固定区間內事件发生的数目. 又設 N_k 是恰有 k 个事件发生的次数，則

$$N_0+N_1+N_2+\cdots=N.\tag{6.4}$$
于是,在 N 次試驗中所观察到的点的总数是
$$N_1+2N_2+3N_3+\cdots=T,\tag{6.5}$$
并且其平均数是 T/N. 如果 N 很大,則我們可期望
$$N_k\approx Np(k;\lambda t)\tag{6.6}$$
(这是概率論所有应用的基础,我們将于第十章中用大数定律給予証明且加以精确化). 将 (6.6) 的結果代入 (6.5), 我們得到
$$\begin{aligned}T&\approx N\{p(1;\lambda t)+2p(2;\lambda t)+3p(3;\lambda t)+\cdots\}\\&=Ne^{-\lambda t}\lambda t\left\{1+\frac{\lambda t}{1}+\frac{(\lambda t)^2}{2!}+\cdots\right\}=N\lambda t,\end{aligned}\tag{6.7}$$

于是

$$\lambda t\approx\frac{T}{N}.\tag{6.8}$$

这关系式提供了一个从观察值去估計 λ 以及将理論与試驗相比較的方法. 我們将在下一节的例中說明这点.

空間中的分布. 我們已考虑过随机事件或随机点在 t 軸上的分布間题，但同样推理可以应用到平面或空間上的点的分布. 我們以面积或体积是 t 的区域来代替长度是 t 的区間,并且,我們的基本假設是任何固定的区域中恰有 k 点发生的概率仅与区域的面积或体积有关而与其形状无关. 此外，我們还有与以前相同的假設: 1) 当 t 小时,在体积为 t 的区域中有多于一点发生的概率与 t 相比較而言是小的; 2) 不相交的区域是彼此独立的，为了求出在体积为 t 的区域中恰有 k 点发生的概率，我們将区域分成許

多子区域．并用 n 次試驗中恰有 k 次成功的概率来逼近我們所求的概率． 这就是說忽略了同一子区域中有多于一点发生的可能性，但由假設 1)，当 $n\to\infty$ 时，其誤差趋于零． 于是，在取极限以后，我們又得到普阿松分布 (6.2)． 天空中的星体，蛋糕中的葡萄干、牧草种子中的杂草种子，物质中的瑕疵，以及原野中动物巢穴等都按照普阿松定律来分布．至于实例，可参看例 (7.b) 与 (7.e)．

7. 符合普阿松分布的观察結果

(a) 放射性分裂． 設放射性物质放射出 α 质点，在时間 t 內到达指定区域的质点数，是随机事件遵从普阿松定律的著名例子．当然，由于放射性物质在不断地縮小，时間太长的話，α 质点的密度就要下降． 但是，就鐳而言，不經过許多年，物质的減少是覺察不出来的． 于是，对于較短的时間区間，情况可考虑为不变的． 因此，这現象滿足导出普阿松分布的理想假設．

在一个著名的放射性物质的試驗中，我們观察 $N = 2608$ 次，而时間区間是 7.5 秒，并且每次都記录下到达指定区域的质点数．表 3 給出有 k 个质点的区間数 N_k，而质点总数是 $T = \Sigma kN_k =$

表 3． 例(a)：放射性分解

k	N_k	$N_p(k; 3.870)$
0	57	54.399
1	203	210.523
2	383	407.361
3	525	525.496
4	532	508.418
5	408	393.515
6	273	253.817
7	139	140.325
8	45	67.882
9	27	29.189
$k\geqslant 10$	16	17.075
总　数	2608	2608.000

10,094，平均数是 $T/N = 3.870$. 理論数值 $N_p(k; 3.870)$ 是很靠近于观察数值 N_k 的. 要判断两者的拟合程度，我們需要对随机起伏的可能范围进行估計. 統計学家是用 χ^2-检驗来判断拟合度的. 由这准則，人們可預期在理想的条件下，100 个类似的情形中，大約有 17 个的拟合程度比表 3 还要差.

(b) 投在伦敦的飞弹. 作为随机点的空間分布的例子，我們考虑第二次世界大战中投在伦敦南部的飞弹的統計. 将整个面积分为 $N = 576$ 块小面积，每小块面积是 1/4 平方公里. 表 4 記录了中 k 个飞弹的面积数[1]. 投下飞弹的总数是 $T = \Sigma k N_k = 537$，平均数是 $\lambda t = T/N = 0.9323\cdots$. 这观察結果对普阿松分布的拟合程度是很好的. 由 χ^2-检驗可知，在理想的条件下，100 次比較中大約有 88 次的拟合度比这观察結果的拟合度还要差. 我們有兴趣的是，大多数人都相信受弹点有聚結的傾向. 果眞如此，中弹很多的地区与沒有中弹的地区岂不是要多些，而中間的情形要少些呢？然而表 4 却明白地表出完善的随机性与均匀性. 这个例証說明了一个公認的事实：随机性到了沒有訓練的眼睛里，就会成为有規則性或有聚結的傾向.

表 4. 例 (b)：投在伦敦的飞弹

k	0	1	2	3	4	5 以上
N_k	229	211	93	35	7	1
$p(k; 0.9323)$	226.74	211.39	98.54	30.62	7.14	1.57

(c) 細胞中染色体的交換. X 射綫的照射产生了有机細胞中的某种变化过程，这种过程叫作染色体的交換. 在連續照射的过程中，发生交換的概率是保持不变的. 按照理論，恰有 k 个交換发生的細胞数 N_k 服从普阿松分布，而且理論上还可預言参数 λ 与

1) 这数字是由文章[46]得出的.

照射強度、温度等等的关系. 但是，我們不打算詳細討論这些內容. 表 5 記录了 11 种不同試驗的結果. 它們是按照拟合度排列的. 最后一列給出了理想情形的逼近的百分比，在这些情形中，随机起伏使得拟合度更坏（用 χ^2–标准检驗）. 发現拟合度在理論值与观察值之間.

表 5. 例(c)：由于 X 射綫的照射而引起的染色体的交换

实驗数		有 k 个交換的細胞				总数 N	χ^2 水准的百分比
		0	1	2	$\geqslant 3$		
1	观察值 N_k $Np(k;0.35508)$	753 752.3	266 267.1	49 47.4	5 6.2	1073	95
2	观察值 N_k $Np(k;0.45601)$	434 432.3	195 197.1	44 44.9	9 7.7	682	85
3	观察值 N_k $Np(k;0.27717)$	280 278.9	75 77.3	12 10.7	1 1.1	368	65
4	观察值 N_k $Np(k;0.11808)$	2278 2280.2	273 269.2	15 15.9	0 0.7	2566	65
5	观察值 N_k $Np(k;0.25296)$	593 589.4	143 149.1	20 18.8	3 1.7	759	45
6	观察值 N_k $Np(k;0.21059)$	639 642.4	141 135.3	13 14.2	0 1.1	793	45
7	观察值 N_k $Np(k;0.28631)$	359 362.0	109 103.6	13 14.9	1 1.5	482	40
8	观察值 N_k $Np(k;0.35372)$	493 498.2	176 167.3	26 28.1	2 3.4	697	35
9	观察值 N_k $Np(k;0.39867)$	793 804.8	339 320.8	62 64.0	5 9.4	1199	20
10	观察值 N_k $Np(k;0.40544)$	579 588.7	254 238.7	47 48.4	3 7.2	883	20
11	观察值 N_k $Np(k;0.49339)$	444 461.6	252 227.7	59 56.2	1 10.5	756	5

(d) 电話接錯. 表 6 是电話接錯的一个統計[1)]，其中被观察的电話叫次共有 $N=267$ 个，而 N_k 表示恰有 k 次接錯的电話数.

---

1) 这統計表是由論文[47]中得来的，这論文含有 32 种不同統計表的图形分析.

我們再一次看到，普阿松分布 $p(k;\ 8.74)$ 对这观察結果的拟合程度是很好的（由 χ^2–检驗，偏差靠近于中值）。在論文[47]中，讀者

表 6. 例 (d)：电話接錯

k	N_k	$Np(k;\ 8.74)$
0–2	1	2.05
3	5	4.76
4	11	10.39
5	14	18.16
6	22	26.45
7	43	33.03
8	31	36.09
9	40	35.04
10	35	30.63
11	20	24.34
12	18	17.72
13	12	11.92
14	7	7.44
15	6	4.33
⩾16	2	4.65
	267	267.00

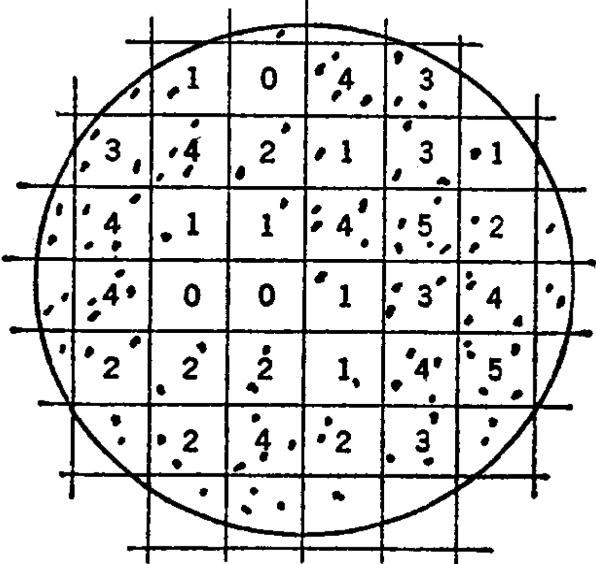


图 1. 細菌在标本片上。

可以找到遵从普阿松定律的其他电話統計. 在某些情形(例如有分机的团体綫及公用电話等)事件之間是有明显的依賴关系的,此时,普阿松分布就不能适合所观察的結果了.

(e) 数細菌与数血球. 图 1 表示涂上細菌羣的 Petri 片的照象,在显微鏡下观看細菌羣呈現为黑点. 将片分成一些小方块. 对于八种不同的細菌作試驗,表 7 記录了恰有 k 个黑点的小方块数[48]. 这里,我們得到了普阿松分布应用于随机点的空間分布的一个重要实际典型的例子.

表 7. 例(e): 細菌的分布

k	0	1	2	3	4	5	6	7	χ^2-水平
观察值 N_k 普阿松理論值	5 6.1	19 18.0	26 26.7	26 26.4	21 19.6	13 11.7	8 9.5		97
同　上	26 27.5	40 42.2	38 32.5	17 16.7	7 9.1				66
同　上	59 55.6	86 82.2	49 60.8	30 30.0	20 15.4				26
同　上	83 75.0	134 144.5	135 139.4	101 89.7	40 43.3	16 16.7	7 7.4		63
同　上	8 6.8	16 16.2	18 19.2	15 15.1	9 9.0	7 6.7			97
同　上	7 3.9	11 10.4	11 13.7	11 12.0	7 7.9	8 7.1			53
同　上	3 2.1	7 8.2	14 15.8	21 20.2	20 19.5	19 15	7 9.6	9 9.6	85
同　上	60 62.6	80 75.8	45 45.8	16 18.5	9 7.3				78

每一行的最后一个数据都包含了更高分組的图形,因而应該标以"k 或者更多个".

8. 等待时間. 負二項分布

考虑 n 次相継的伯努利試驗,我們問: 一直到出現第 r 次成功为止到底需要作多少次試驗. 其中 r 是一个固定的正整数. 当然, n 次試驗的成功次数可以少于 r,但第 r 次成功发生在第

$\nu \leqslant n$ 次試驗的概率不依賴于 n，而只依賴于 ν，r 和 p. 因为一定有 $\nu \geqslant r$，所以我們愿意把 ν 写作 $\nu = r + k$. 第 r 次成功发生在第 $r + k$ 次試驗（其中 $k = 0, 1, \cdots$）的概率用 $f(k; r, p)$ 来表示. 它等于在 r 次成功之前恰有 k 次失敗的概率. 而这个事件的发生，当且仅当 $r + k - 1$ 次試驗中恰有 k 次失敗而下一次（即第 $r + k$ 次）一定出現成功；但它們的概率分别为 $\binom{r+k-1}{k} \times p^{r-1}q^k$ 和 p，所以

$$f(k; r, p) = \binom{r+k-1}{k} \cdot p^r q^k. \tag{8.1}$$

用第二章 (12.4) 把二項式系数改写一下，便得另一形式：

$$f(k; r, p) = \binom{-r}{k} p^r (-q)^k, \quad k = 0, 1, 2, \cdots. \tag{8.2}$$

現在假定伯努利試驗一直繼續到第 r 次成功出現为止. 一个标准的样本点可以用这样一个序列来表示：具有 k 个 F（k 任意）恰有 r 个 S 而且序列的最后字母是 S；由定义，这样的点的概率为 $p^r q^k$. 然而，我們必須問：試驗一直作下去不完的可能是否存在，也就是說，是否有无穷試驗序列，其成功次数少于 r. 因为 $\sum_{k=0}^{\infty} f(k; r, p)$ 是第 r 次成功发生在有限次試驗的概率；因此，一个无穷試驗序列中所包含的成功次数少于 r 这个事件不可能发生当且仅当

$$\sum_{k=0}^{\infty} f(k; r, p) = 1. \tag{8.3}$$

为了証明 (8.3) 成立，只須注意：由二項定理

$$\sum_{k=0}^{\infty} \binom{-r}{k} (-q)^k = (1-q)^{-r} = p^{-r}. \tag{8.4}$$

将 (8.4) 乘以 p^r 我們得到 (8.3).

在我們的等待时間問題中，r 必須是固定的正整数，但是由 (8.1) 或 (8.2) 所定义的量是非負的，而且(8.3)对任何正数 r 都成立. 对于任何一个固定的实数 $r > 0$ 和 $0 < p < 1$，序列$\{f(k; r, p)\}$

叫作負二項分布. 在很多应用中会碰到这一分布.(在第五章問題 24 里它是作为波利亚分布的极限形式出現的.)当 r 为正整数时,$\{f(k; r, p)\}$ 可以解释为第 r 次成功所需之等待时間的概率分布;正由于此,我們也把它叫作巴斯加(Pascal)分布. 当 $r=1$ 时,它化为几何分布 $\{pq^k\}$.

表 8. 概率(8.5)

r	u_r	U_r	r	u_r	U_r
0	0.079 589	0.079 589	15	0.023 171	0.917 941
1	.079 589	.159 178	16	.019 081	.937 022
2	.078 785	.237 963	17	.015 447	.952 469
3	.077 177	.315 140	18	.012 283	.964 752
4	.074 790	.389 931	19	.009 587	.974 338
5	.071 674	.461 605	20	.007 338	.981 676
6	.067 902	.529 506	21	.005 504	.987 180
7	.063 568	.593 073	22	.004 041	.991 220
8	.058 783	.651 855	23	.002 901	.994 121
9	.053 671	.705 527	24	.002 034	.996 155
10	.048 363	.753 890	25	.001 392	.997 547
11	.042 989	.796 879	26	.000 928	.998 475
12	.037 676	.834 555	27	.000 602	.999 077
13	.032 538	.867 094	28	.000 379	.999 456
14	.027 676	.894 770	29	.000 232	.999 688

假定开始时每盒火柴有 50 根,u_r 是发現第一盒空时第二盒恰有 r 根火柴的概率. $U_r = u_0 + u_1 + \cdots + u_r$ 是对应第二盒不多于 r 根的概率.

例. 巴拿赫火柴盒問題. 某一数学家經常在左边口袋放一盒火柴,右边口袋放一盒火柴. 当他要用火柴的时候,它随机地从一个口袋中去取火柴,因此連續的抽取构成了一串 $p=\frac{1}{2}$ 的伯努利試驗. 假定最初每盒火柴恰巧包含 N 根,我們考慮:数学家第一次发現空盒子的时刻. 在那一时刻,另一盒火柴可能还有 $0, 1, \cdots, N$ 根火柴,我們用 u_r 表示其对应的概率. 我們令从左边口

袋选取为“成功”. “当发現左边口袋的火柴盒空时右边的火柴盒中尚有 r 根火柴”这一事件当且仅当下列事件发生时才发生：恰有 $N-r$ 次失敗发生在第 $N+1$ 次成功之前. 这个事件的概率为 $f\left(N-r;N+1,\frac{1}{2}\right)$. 同样的推理可以用之于右边口袋. 因此，所要求的概率为

$$u_r = 2f\left(N-r;N+1,\frac{1}{2}\right) = \binom{2N-r}{N}2^{-2N+r}. \tag{8.5}$$

表 8 給出了 $N=50$ 时 u_r 的数値 [参看問題 21—23 和第九章例 (3.f)].

9. 多項分布

二項分布很容易推广到一般的情形，即每次試驗有若干个可能結果的 n 次重复独立試驗. 設 $E_1,\cdots,E_r$ 是每次試驗的可能結果，且每次試驗中 E_i 发生的概率是 $p_i(i=1,2,\cdots,r)$. 当 $r=2$ 时，它化为伯努利試驗序列；在一般情形，p_i 应满足以下的条件

$$p_1+p_2+\cdots+p_r=1 \qquad p_i \geqslant 0. \tag{9.1}$$

n 次試驗的可能結果可表为形如 $E_3E_1E_2\cdots$ n 元序列. 并且在 n 次試驗中 E_1 出現 k_1 次，E_2 出現 k_2 次，$\cdots$，E_r 出現 k_r 次的概率是

$$\frac{n!}{k_1!\,k_2!\cdots k_r!}\,p_1^{k_1}p_2^{k_2}p_3^{k_3}\cdots p_r^{k_r}, \tag{9.2}$$

其中 k_i 是任何非負的整数并且滿足

$$k_1+k_2+\cdots+k_r=n. \tag{9.3}$$

当 $r=2$ 时，(9.2) 就化为二項分布，其中 $p_1=p$，$p_2=q$，$k_1=k$，$k_2=n-k$. 从第二章公式 (4.7) 出发，我們可用証明二項分布的方法来証明这个一般情形的公式.

因公式(9.2)是 $(p_1+\cdots+p_r)^n$ 多項展开式中的一般項，故 (9.2) 称为多項分布. 这分布主要应用于个体被区分为多于两类

的有放回抽样(例如,按照职业进行分类).

例. (a) 擲 12 个骰子，問骰子的每面都出現两次的概率是多少? 設 $E_1, E_2, \cdots, E_6$ 表示骰子的六个面，由于 $k_i = 2, p_i = \frac{1}{6}, i = 1, 2, \cdots, 6$, 所以我們的答案是 $(12!)(2)^{-6}(6)^{-12} = 0.0034\cdots$.

(b) 抽样. 把一个具有N个元素的总体分为 r 个子类 $E_1, \cdots, E_r$, 子类的大小分别为 $Np_1, \cdots, Np_r$. 多項分布給出了由这个总体中有放回地抽取一个大小为 n 的随机样本的各种可能組合的概率.

(c) 多重伯努利試驗. 两串伯努利試驗合在一起可考虑为一个具有四个可能結果 $(S, S), (S, F), (F, S), (F, F)$ 的联合試驗,其中 p_1, q_1 与 p_2, q_2 分别为这两串伯努利試驗的成功概率与失敗概率. 如果这两串試驗是独立的, 则联合試驗中四个可能結果的概率分别为 $p_1p_2, p_1q_2, q_1p_2, q_1q_2$. 若 k_1, k_2, k_3, k_4 是四个整数,其和为 n, 则 n 次联合試驗中 SS 出現 k_1 次, SF 出現 k_2 次, FS 出現 k_3 次, FF 出現 k_4 次的概率是

$$\frac{n!}{k_1!\,k_2!\,k_3!\,k_4!}\, p_1^{k_1+k_2} q_1^{k_3+k_4} p_2^{k_1+k_3} q_2^{k_2+k_4}. \tag{9.4}$$

抽样检查过程可作为我們的一个特例. 一个成品是合格或废品的概率分别为 p, q, 成品是否被检查的概率分别为 p', q', 因成品是否被检查与它的質量无关, 所以我們有两串独立試驗 [参看問題 25 与 26, 以及第九章問題 12].

10. 問 題

1. 設性別的分布都是等可能的,問六个小孩的家庭中恰有三男三女的概率是多少?

2. 一个玩桥牌者連續三次沒有得到爱司,他是否有理由埋怨运气不好?

3. 随机数字序列要有多长才能使得序列中数字 7 出現的概率至少是 9/10?

4. 設在玩桥牌中每次桥牌分配是独立的,問需玩多少次才能使得一个事

先指定的玩牌者至少有一次得四张爱司的概率大于 1/2? 如果"将事先指定的玩牌者"换为"有一个玩牌者",則結論又如何?

5.設中靶的概率是 1/5, 問十次独立的打靶中至少有两次中靶的概率是多少?

6.在上題中, 如果已知至少有一次中靶, 問至少两次中靶的条件概率是多少?

7.一組任意选取的 13 张桥牌中, 恰有两张紅牌的概率是多少? 并将这概率与成功概率为 $p=\frac{1}{2}$ 的伯努利試驗中的相应概率相比較 [关于桥牌的描述, 可参看第一章的脚注].

8.六个人的生日都在十二个月中的任意两个月份, 而在其它的十个月份內沒有这六个人的生日的概率是多少? [假設每个人的生日是彼此独立的, 并設, 生日发生在某个月份是等可能的.]

9.擲六个骰子, 求以下的概率: (a) 至少有一个么点出現; (b) 恰有一个么点出現; (c) 恰有两个么点出現. 并与相应的普阿松近似值相比較.

10.如果慣用左手者的平均百分数是 1%, 試計算 200 人中至少有四个为慣用左手者的概率.

11.設一本 500 頁的书中含有 500 个錯处, 試計算一頁中至少有三个錯处的概率.

12.如果总体中色盲者的百分数是 1%, 問有放回的随机样本应該多大才能使样本中有一个色盲者的概率大于或等于 0.95?

13.在上題中, 如果样本的大小是 100, 求以下的概率: (a) 样本中沒有色盲者; (b) 样本中至少有两个色盲者.

14.如果想让蛋糕中至少含有一粒葡萄干的概率大于或等于 0.99, 試計算蛋糕中葡萄干的平均数.

15.扑克游戏中出現最大同花的概率是 $p=1/649,740$, 問 n 要多大才能使 n 次扑克游戏中沒有最大同花出現的概率小于 $\frac{1}{e}\approx\frac{1}{3}$ (注: 不用計算便可得出答案).

16.一本书包含 n 頁, 每一頁的印錯处平均为 λ. 計算至少有一頁其中印錯之处超过 k 的概率.

17.假定有两类顆粒(例如点心中的不同类的葡萄干, 質料中的不同类的杂質). 在一块給定的体积中, 第一类的顆粒恰有 i 个的概率为 $p(i;a)$, 第

二类的顆粒恰有 k 个的概率为 $p(k; b)$，而且这两个事件是假定为相互独立的. 証明这个給定的体积中总共包含 n 个顆粒的概率为 $p(n; a+b)$（把这一論断和假設加以抽象化）.

18. 交通問題. 街道交叉口上的交通的暢通程度是用任意給定的一秒鐘內有一部車子通过的概率 p 来描述的；而且在不同秒的时刻車子的經过是互不影响的. 把秒作为不可分的时刻单位，于是伯努利試驗模型可以应用. 假定一个步行者只能在后三秒鐘沒有車子通过的情形下才能跨过街道. 求出一个步行者恰巧等待 $k=0, 1, 2, 3, 4$ 秒的概率.（其一般的公式并不很显然，在第十三章第 7 节討論成功的連貫理論时将把它推出来.）

19. 設两个人都扔 n 次錢币. 求錢币出現正面的次数相等的概率.

20. 在一串成功概率为 p 的伯努利試驗序列中，求 a 个成功出現在 b 个失敗以前的概率.（注：这个結果至多在 $a+b-1$ 次試驗以后就可决定. 这問題在賭博的古典理論中起着作用，它和下面的如何分賭注的問題有关：当一方差 a 点获全胜，另一方差 b 点获全胜时賭博中断，如何分配賭注.）

21. 在巴拿赫火柴盒問題（第 8 节）中，求出：一盒恰用完（而是拿出来发現它是空的）时另一盒尙恰有 r（其中 $r=1, 2, \cdots, N$）根火柴的概率.

22（續上）. 应用上面的結果，求出首先用完的那一盒不是最初拿出而发現它是空的的概率 x. 并証明由此所得之表达式可化为 $x=\binom{2N}{N}2^{-2N}$，或者近似地为 $(N\pi)^{-1/2}$.

23. 两个校对者独立地校对某一本书的印样，他們各发現 k_1 和 k_2 个錯誤，其中有 k_{12} 个錯誤被这两个人同时发現. 試給出印样的錯誤的个数 n 的一个合理的估計.（假定这两个校对者对应于伯努利試驗，他們发現一个錯誤的概率分别为 p_1 和 p_2. 可应用大数定律.）

注意：这个問題是用簡单术語来描述罗泽福特（Rutherford）的，閃爍計数試驗装备的.

24. 用陷井来估計一羣野兽的总体的多少[49]，接連設立 r 次陷井. 假定每一个野兽陷入的概率都同样为 q；最初共有 n 个野兽；只当野兽被捕（拿走）时，两次相継設立陷井的情况才会改变. 求出 r 次陷井捕获的野兽分别为 $n_1, n_2, \cdots, n_r$ 的概率.

25. 多維伯努利試驗. 在例（9.c）中，如果知道 (S, F) 与 (F, S) 中有一个发生，分别求 (S, F)，(F, S) 发生的条件概率 p 及 q. 証明当 $p_1>p_2$ 时，

$p>\frac{1}{2}$; 当 $p_2>p_1$ 时，$p<\frac{1}{2}$.

26(續上)[1]. 如果已知在 n 对試驗中 (S, F) 与 (F, S) 共发生 m 次，証明 (S, F) 恰巧发生 k 次的概率是 $b(k; n, p)$.

27. 二項分布与普阿松分布的結合. 設一只昆虫生 r 个卵的概率是 $p(r; \lambda)$，而卵能发育为成虫的概率是 p. 又設每个卵是否发育为成虫是彼此独立的. 証明有 k 个后代的概率是遵从参数为 λp 的普阿松分布.

注意：同样情况下的另一个例子：k 个染色体分裂的概率为 $p(k; \lambda)$，一个分裂的染色体恢复原状的概率为 p.（类似的一些其它的例子，請参看第九章例 (1.d) 及第十二章第 1 节.）

28. 証明定理[2]：多項分布 (9.2) 的最大項滿足不等式

$$np_i - 1 < k_i \leqslant (n + r - 1)p_i, \quad i = 1, 2, \cdots, r. \tag{10.1}$$

（提示：首先証明，最大項的充分必要条件是 $p_i k_j \leqslant p_j(k_i + 1)$（对每一对 (i, j)）. 把这些不等式对全部的 j 加起来，同时也对全部 $i \neq j$ 加起来.）

29. 当 k 是不超过 λ 的最大整数时，普阿松分布中的項 $p(k; \lambda)$ 达到最大值.

注意：問題 30—34 是关于二項分布的普阿松逼近的. 把 np 了解为 λ，m 了解为不超过 $(n + 1)p$ 的最大整数（也就是說，m 是二項分布的中心項的指标）.

30. 証明当 k 从 0 跑到 ∞ 时，$a_k = b(k; n, p)/p(k; \lambda)$ 最初上升而后下降，而且在 $k = m$ 时达到它的最大值.

31. 当 k 增大时，$b(k; n, p)$ 最初比 $p(k; \lambda)$ 小而后变大最后又变得比 $p(k; \lambda)$ 小.

32. 如果 $n \to \infty$，$p \to 0$ 使得 $np = \lambda$ 保持为常数，则

$$b(k; n, p) \to p(k; \lambda)$$

对全部 k 一致地成立.

33. 証明：

---

1) 参看 [50]. 瓦尔德 (Wald) 利用这个結果設計了一个实用的方法，用于比較两串由經驗給定的試驗（例如，两架机器的产品），而选出成功概率較大的一串. 他把这問題簡化为如下的問題，在一串伯努利試驗里求成功頻率是否显著地不同于 1/2.

2) 在第一版中，只断言 $|k_i - np_i| \leqslant r$. 此处的改进和精練的証明都属于莫萊 (P. A. P. Moran).

$$\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^{n-k} \geqslant b(k;n,p) \geqslant \frac{\lambda^k}{k!}\left(1-\frac{k}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^n. \qquad (10.2)$$

34.应用第二章不等式 (8.12)，从 (10.2) 推出

$$p(k;\lambda)e^{k\lambda/n} > b(k;n,p) > p(k;\lambda)e^{-k^2/(n-k)-\lambda^2/(n-\lambda)}. \qquad (10.3)$$

注意：虽然(10.2)非常粗糙，但(10.3)却給出較精确的誤差估計．用类似于第二章第 9 节的計算方法可以很容易地改进 (10.3)．附带提一下：应用問題 30 的結果，容易看出 (10.3) 左边的指数可以用 $m\lambda/n$ 来代替，而后者 $\leqslant(p+n^{-1})\lambda$.

其他的极限定理

35.超几何分布的二項逼近．一个包含有 N 个元素的总体分为紅黑两类元素，其比例为 $p:q$（其中 $p+q=1$）．无放回地抽取一个大小为 n 的样本．其中恰有 k 个紅元素的概率由第二章第 6 节的超几何分布所給出．証明：当 $N\to\infty$ 时这个概率趋于 $b(k;n,p)$.

36.在前面一个問題中，令 p 很小 n 很大而 $\lambda=np$ 大小适当．于是超几何分布可以用普阿松分布 $p(k;\lambda)$ 来逼近．不用二項逼近来直接驗証这个关系．

37.在第 8 节的負二項分布 $\{f(k;r,p)\}$ 中，令 $q\to 0$ 和 $r\to\infty$ 使 $rq=\lambda$ 保持常数．証明：

$$f(k;r,p)\to p(k;\lambda).$$

38.多重的普阿松分布．当 n 很大而 $np_j=\lambda_j$ 大小适当时，多項分布 (9.2)可以用

$$e^{-(\lambda_1+\cdots+\lambda_r)}\frac{\lambda_1^{k_1}\lambda_2^{k_2}\cdots\lambda_r^{k_r}}{k_1!\,k_2!\cdots k_r!}$$

来逼近．并証明这些項之和为 1．（注意：問題 17 和二重普阿松分布有联系．）

---

39. (a) 应用显然的关系

$$b(k;n,p)=b(n-k;n,q)$$

从 (3.5) 直接推出 (3.6)．(b) 用归納法和从第四章 (3.1) 的一般的求和公式推出二項分布．

40.証明 $\Sigma kb(k;n,p)=np$ 和 $\Sigma k^2b(k;n,p)=n^2p^2+npq$.

41. 証明 $\Sigma k^2 p(k;\lambda) = \lambda^2 + \lambda$.

42. 驗証等式

$$\sum_{\nu=0}^{k} b(\nu; n_1, p) b(k-\nu; n_2, p) = b(k; n_1+n_2, p), \qquad (10.4)$$

并解释这等式的概率意义. 提示: 利用第二章的公式(6.4).

注意: 等式 (10.4) 是卷积的特殊情形,这在第十一章将提到,而另一个特例是(10.5).

43. 驗証等式

$$\sum_{\nu=0}^{k} p(\nu; \lambda_1) p(k-\nu; \lambda_2) = p(k; \lambda_1+\lambda_2). \qquad (10.5)$$

44. 設

$$B(k; n, p) = \sum_{\nu=0}^{k} b(\nu; n, p) \qquad (10.6)$$

是 n 次試驗中至多有 k 个成功的概率,則

$$B(k; n+1, p) = B(k; n, p) - pb(k; n, p), \qquad (10.7)$$
$$B(k+1; n+1, p) = B(k; n, p) + qb(k+1; n, p).$$

試給 (a) 由定义出发的; (b) 分析的,証明.

45. 証明

$$B(k; n, p) = (n-k)\binom{n}{k}\int_0^q t^{n-k-1}(1-t)^k dt \qquad (10.8)$$

与

$$1 - B(k; n, p) = n\binom{n-1}{k}\int_0^p t^k (1-t)^{n-k-1} dt. \qquad (10.9)$$

提示: 用分部积分法或将等式的两边对 p 求导数. 这两个結果可以互相推导.

注意: (10.9)中的积分是不完全的貝塔 (Beta) 函数. 对于 k 与 n 从 0 变到 50,以及 $p = 0.01, 0.02, 0.03, \cdots, 1 - B(k; n, p)$ 的七位小数表是由下面的书: "K. Pearson, Table of incomplete beta function, London (Biometrika Office), 1934" 給出.

46. 証明:

$$p(0;\lambda) + \cdots + p(n;\lambda) = \frac{1}{n!}\int_\lambda^\infty e^{-x} x^n dx. \qquad (10.10)$$

注意: 在下面的問題中我們对二項分布所有各項給以一个上界. 計算是很簡单的, 其方法經修改后可以給出德模佛-拉普拉斯极限定理一个最簡

单的推导(参看第七章問题 19—21)。令

$$\xi_k = \frac{k-(n+1)p+\frac{1}{2}}{\{(n+1)pq\}^{1/2}}, \tag{10.11}$$

且令 m 为中心項的指标;即是說,m 是滿足(3.2)的整数。

47.証明当 $r \geqslant (n+1)p$ 时有

$$b(r;n,p) \leqslant b(m;n,p)e^{-\frac{1}{2}p\xi_r^2+\delta^2}, \tag{10.12}$$

其中 $\delta = m-(n+1)p+\frac{1}{2}$, 从而 $|\delta|<\frac{1}{2}$.

提示: 把(3.1)写成下述形式:

$$\frac{b(k;n,p)}{b(k-1;n,p)} = \frac{(n+1)pq-\{k-(n+1)p\}p}{(n+1)pq+\{k-(n+1)p\}q}. \tag{10.13}$$

对 $k \geqslant (n+1)p$ 可推出

$$\log\frac{b(k;n,p)}{b(k-1;n,p)} < \log\left\{1-\frac{k-(n+1)p}{(n+1)pq}p\right\} <$$

$$< -\frac{k-(n+1)p}{(n+1)pq}p, \tag{10.14}$$

从而求和即得所需之結論.

48.对于 $r \leqslant (n+1)p$, 不等式(10.12)仍然成立,只不过把指数上的 p 换成 q 罢了. 因此,如果以 pq 代 p, 則不等式对全部 r 都成立.

第七章　二項分布的正态逼近

1. 正 态 分 布

为了以后的方便，我們在这里先引进两个重要的函数。

定义． 函数

$$\phi(x) = \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}x^2} \tag{1.1}$$

称为正态密度函数；它的积分

$$\Phi(x) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{x} e^{-\frac{1}{2}y^2}\, dy \tag{1.2}$$

就叫作正态分布函数．

$\phi(x)$ 的图形是对称的鈡形曲綫（如图 1 所示）．但須注意，图中两个坐标軸所取的单位是不同的． $\phi(x)$ 的最大值为 $(2\pi)^{-\frac{1}{2}} \approx 0.399$． 所以，在通常的（同单位的）笛卡尔坐标里 $y = \phi(x)$ 的图形比画出来的要扁平得多．

引理 1． $\phi(x)$ 的图形与 x 軸所围成的区域的面积为 1，即是

$$\int_{-\infty}^{\infty} \phi(x)\, dx = 1. \tag{1.3}$$

証． 我們有

$$\begin{aligned} \left\{\int_{-\infty}^{\infty} \phi(x)\, dx\right\}^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi(x)\phi(y)\,dx\,dy \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\,dx\,dy. \end{aligned} \tag{1.4}$$

这个重积分可用极坐标来表达：

$$\frac{1}{2\pi}\int_0^{2\pi} d\theta \int_0^{\infty} e^{-\frac{1}{2}r^2} r\, dr = \int_0^{\infty} e^{-\frac{1}{2}r^2} r\, dr = -e^{-\frac{1}{2}r^2}\Big|_0^{\infty} = 1, \tag{1.5}$$

至此，引理証毕．

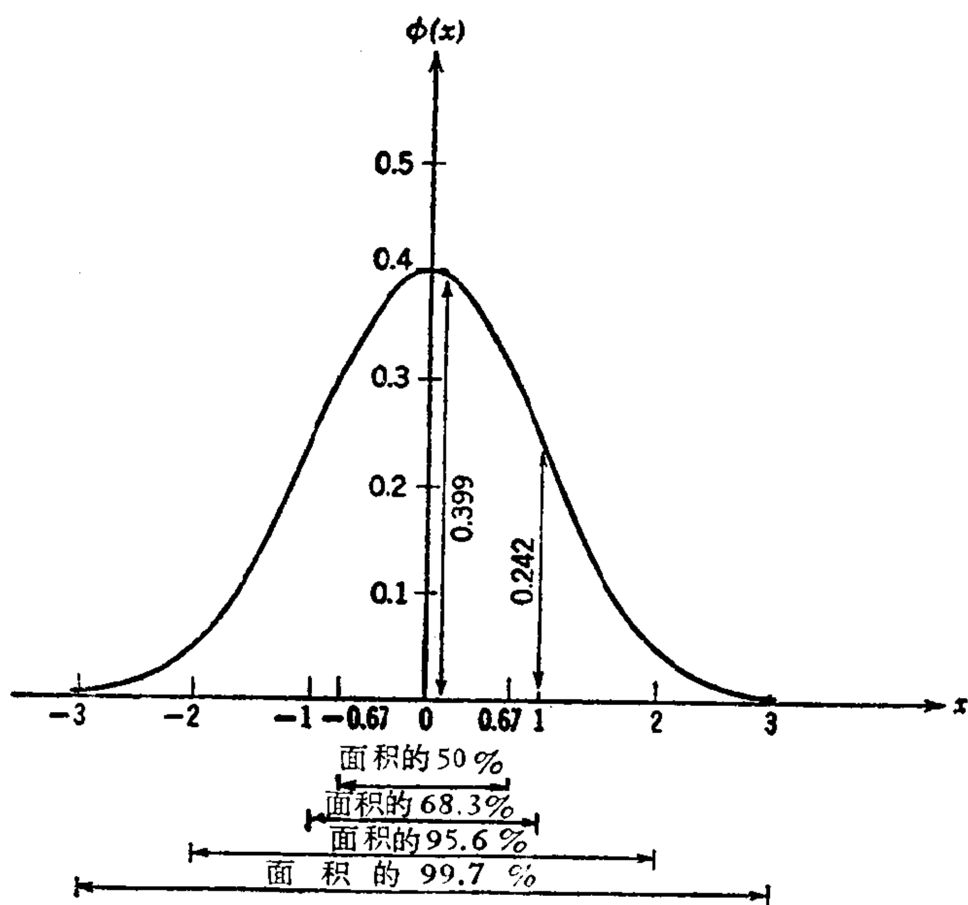


图 1. 正态密度函数.

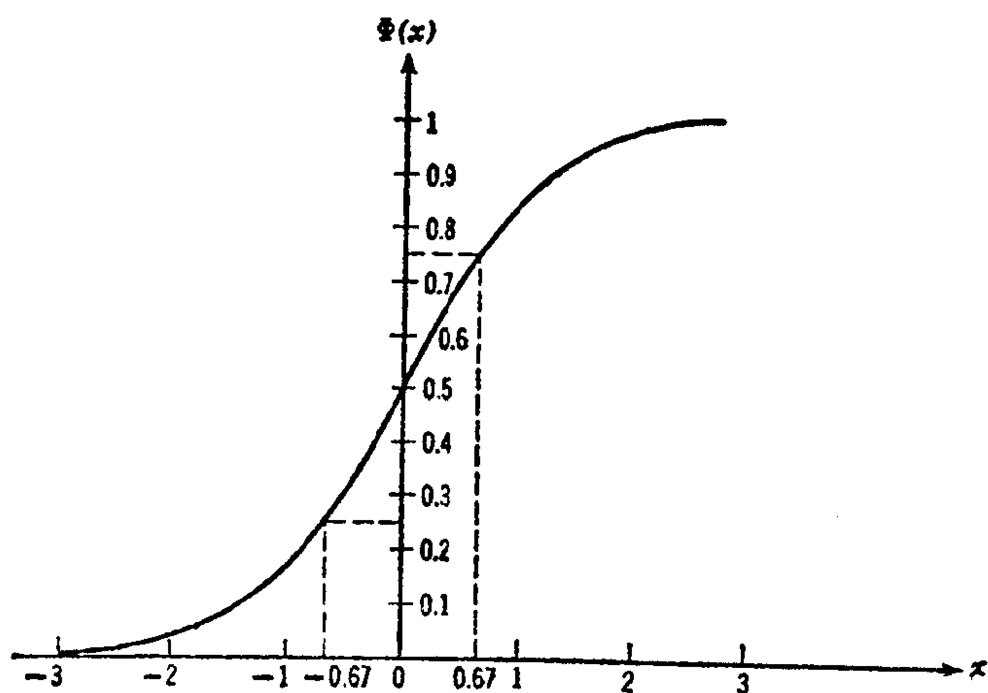


图 2. 正态分布函数.

从定义及引理可推出：$\Phi(x)$ 由 0 单調上升到 1，它的图形是一个 S 形的曲綫（如图 2 所示），且有

$$\Phi(-x) = 1 - \Phi(x). \tag{1.6}$$

表 1 給出当 x 取正数时 $\Phi(x)$ 的函数值[1]，而从 (1.6) 又可得出 $\Phi(-x)$.

在很多情况下，当 x 很大时，对“尾量” $1 - \Phi(x)$ 有一个初等的估計是合宜的．下面将給出这个估計．

引理 2[2]. 当 $x \to \infty$ 时

$$1 - \Phi(x) \sim \frac{1}{(2\pi)^{\frac{1}{2}} x} e^{-\frac{1}{2}x^2} \tag{1.7}$$

更精确些說，对任意 $x > 0$，下面的双重不等式

$$\frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}x^2} \left\{\frac{1}{x} - \frac{1}{x^3}\right\} < 1 - \Phi(x) < \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}x^2} \cdot \frac{1}{x} \tag{1.8}$$

成立（参看問題 1）.

証． 用微分法我們可以驗証

$$\frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}x^2} \frac{1}{x} = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_x^\infty e^{-\frac{1}{2}y^2} \left\{1 + \frac{1}{y^2}\right\} dy. \tag{1.9}$$

(1.9)右边的被积函数大于

$$1 - \Phi(x) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_x^\infty e^{-\frac{1}{2}y^2}\, dy \tag{1.10}$$

右边的被积函数．这就証明了 (1.8) 中的第二个不等式．用同样的办法可証 (1.8) 的第一个不等式，只要另取一个新的被积函数 $e^{-\frac{1}{2}y^2}\left\{1 - \frac{3}{y^4}\right\}$，其中 $e^{-\frac{1}{2}y^2}\left\{1 - \frac{3}{y^4}\right\}$ 小于 $e^{-\frac{1}{2}y^2}$.

术語注解. 在数学文献中术語分布函数就是指这样的非降函数 $F(x)$：当 $x \to -\infty$ 时 $F(x)$ 趋于 0，而当 $x \to \infty$ 时 $F(x)$ 趋于 1．統計学家把此函

---

1) 至于更大的表，則請参看[51]．在那个表里当 x 从 0 变到 1 时每隔 0.0001 取一个分点，当 $x > 1$ 时則每隔 0.001 取一个分点，在这些分点上給出 $\varphi(x)$ 与 $\Phi(x) - \Phi(-x)$ 的値直到 15 位小数.

2) 此处及以后符号“$\sim$”皆表示两边之比趋于 1.

表 1. 正态分布

t	$\phi(t)$	$\Phi(t)$	t	$\phi(t)$	$\Phi(t)$
0.0	0.398 942	0.500 000	2.5	0.017 528	0.993 790
0.1	0.396 952	0.539 828	2.6	0.013 583	0.995 339
0.2	0.391 043	0.579 260	2.7	0.010 421	0.996 533
0.3	0.381 388	0.617 911	2.8	0.007 915	0.997 445
0.4	0.368 270	0.655 422	2.9	0.005 953	0.998 134
0.5	0.352 065	0.691 462	3.0	0.004 432	0.998 650
0.6	0.333 225	0.725 747	3.1	0.003 267	0.999 032
0.7	0.312 254	0.758 036	3.2	0.002 384	0.999 313
0.8	0.289 692	0.788 145	3.3	0.001 723	0.999 517
0.9	0.266 085	0.815 940	3.4	0.001 232	0.999 663
1.0	0.241 971	0.841 345	3.5	0.000 873	0.999 767
1.1	0.217 852	0.864 334	3.6	0.000 612	0.999 841
1.2	0.194 186	0.884 930	3.7	0.000 425	0.999 892
1.3	0.171 369	0.903 200	3.8	0.000 292	0.999 928
1.4	0.149 727	0.919 243	3.9	0.000 199	0.999 952
1.5	0.129 518	0.933 193	4.0	0.000 134	0.999 968
1.6	0.110 921	0.945 201	4.1	0.000 089	0.999 979
1.7	0.094 049	0.955 435	4.2	0.000 059	0.999 987
1.8	0.078 950	0.964 070	4.3	0.000 039	0.999 991
1.9	0.065 616	0.971 283	4.4	0.000 025	0.999 995
2.0	0.053 991	0.977 250	4.5	0.000 016	0.999 997
2.1	0.043 984	0.982 136			
2.2	0.035 475	0.986 097			
2.3	0.028 327	0.989 276			
2.4	0.022 395	0.991 802			

数常常喜欢叫作“累积分布函数”这一术語. 其实“累积”这一形容詞是多余的. 所謂密度函数就是这样的非負函数: 其在整个 x 軸上的积分为 1. 任何密度函数从 $-\infty$ 到 x 的积分是一个分布函数. 旧术語“頻率函数”是密度函数的同义詞.

正态分布函数常常也叫作“高斯分布”(Gaussian distribution). 但是,在

高斯之先，德模佛和拉普拉斯就在概率論中用过这个函数。如果改变坐标原点与变量单位，則可使 $\Phi(x)$ 变为 $\Phi((x-a)/b)$；后者就叫作具有均值 a 及方差 b^2（或标准差 $|b|$）的正态分布函数。函数 $2\Phi(x2^{1/2})-1$ 常常叫作誤差函数。

2. 德模佛-拉普拉斯极限定理

令 S_n 表示成功概率为 p 的 n 次伯努利試驗中成功的次数，于是 $b(k;n,p)$ 就是事件"$S_n=k$"的概率。实际上，我們感兴趣的往往是关于成功次数落在預定的范围 α 与 β 之間的概率。如果 α、β 都是整数而且 $\alpha<\beta$，則此事件由不等式 $\alpha\leqslant S_n\leqslant\beta$ 所定义。而且它的概率为

$$\mathbf{P}\{\alpha\leqslant S_n\leqslant\beta\}$$
$$=b(\alpha;n,p)+b(\alpha+1;n,p)+\cdots+b(\beta;n,p).\quad(2.1)$$

这个和数可能包括很多項，而直接計算往往是不容易办到的，然而幸运的是：当 N 相当大时，概率(2.1)可用正态分布函数来簡单地逼近。这是德模佛[1]和拉普拉斯[2]所发現的。我們将要看到，它的重要性要远远超出数値計算的范围。

我們的第一个目的是要导出个别項

$$b(k;n,p)=\frac{n!}{k!(n-k)!}p^kq^{n-k}\quad(2.2)$$

的漸近公式。設概率 p 保持不变，而註 $n\to\infty$。根据大数定律[第六章(4.2)]，对于每一个 $\varepsilon>0$，$|\mathbf{S}_n-np|>n\varepsilon$ 的概率都趋于0，因此，只有滿足 $|k-np|n^{-1}\to0$ 的 k 会成問題的。为了方便起見，引进新的变量 $\delta_k=k-np$。于是

$$k=np+\delta_k,\qquad n-k=nq-\delta_k;\quad(2.3)$$

而我們感兴趣的只是 n、k 的这样的組合：$n\to\infty$ 和 $\delta_k/n\to0$。

---

1) Abraham DeMoivre (1667—1754). 他的文章[52]发表于1718.

2) Pierre S. Laplace (1749—1827). 他的文章[53]发表于1812.

把(2.2)中的因子用史特令公式表示，我們得到[1)]

$$b(k;n,p)\sim\left\{\frac{n}{2\pi k(n-k)}\right\}^{\frac{1}{2}}\left(\frac{np}{k}\right)^{k}\left(\frac{nq}{n-k}\right)^{n-k}=$$

$$=\left\{\frac{n}{2\pi(np+\delta_k)(nq-\delta_k)}\right\}^{\frac{1}{2}}\frac{1}{(1+\delta_k/np)^{np+\delta_k}(1-\delta_k/nq)^{nq-\delta_k}},\tag{2.4}$$

其中符号"～"表示两边之比趋于1.

为了計算最后一个分式，可先取其对数. 在区間 $|\delta_k|\leqslant npq$ 內可应用泰劳展式[第二章公式(8.9)]，而其分母的对数为

$$(np+\delta_k)\log(1+\delta_k/np)+(nq-\delta_k)\log(1-\delta_k/nq)=$$

$$=(np+\delta_k)\left(\frac{\delta_k}{np}-\frac{\delta_k^2}{2n^2p^2}+\frac{\delta_k^3}{3n^3p^3}-+\cdots\right)-$$

$$-(nq-\delta_k)\left(\frac{\delta_k}{nq}+\frac{\delta_k^2}{2n^2q^2}+\frac{\delta_k^3}{3n^3q^3}+\cdots\right).\tag{2.5}$$

按 δ_k 的冪次整理一下，我們得

$$\frac{\delta_k^2}{2n}\left(\frac{1}{p}+\frac{1}{q}\right)-\frac{\delta_k^3}{6n^2}\left(\frac{1}{p^2}-\frac{1}{q^2}\right)+\cdots=$$

$$=\frac{\delta_k^2}{2npq}\left\{1+\frac{p-q}{3pq}\cdot\frac{\delta_k}{n}+\cdots\right\},\tag{2.6}$$

其中 $\delta_k^2/2npq$ 是主要項，因为 $\delta_k/n\to 0$. 如果我們假定 $\delta_k^3/n^2\to 0$，則(2.6)中除了第一項外都趋于0，因此(2.4)可写成較簡单的形式

$$b(k;\,n,\,p)\sim\left\{\frac{n}{2\pi(np+\delta_k)\,(nq-\delta_k)}\right\}^{\frac{1}{2}}e^{-\delta_k^2/2npq}.\tag{2.7}$$

然而，$np+\delta_k\sim np$，$nq-\delta_k\sim nq$，因此(2.7)可以更进一步地

---

1) 回忆一下，在第二章中我們并沒有完全証明史特令公式，仅仅証明了：$r!\sim Cr^{r+\frac{1}{2}}e^{-r}$，其中 C 为正的常数. 而在书中已假定了：$C=(2\pi)^{\frac{1}{2}}$. 欲証 $C=(2\pi)^{\frac{1}{2}}$，需把(2.4)，(2.7)，(2.8)中的因子 $(2\pi)^{\frac{1}{2}}$ 用 C 来代替. 这样，在(2.11)，(2.14)和(2.18)的右边就必須要插进 $C/(2\pi)^{\frac{1}{2}}$ 这个因子. 为了証明这个因子等于1，只要取 x_β 和 $x_{-\alpha}$ 都很大，則在修正了的方程(2.18)的右边可任意接近 $C/(2\pi)^{\frac{1}{2}}$，而左边則接近于1（参看第六章第3节）.

簡化为

$$b(k;n,p)\sim\frac{1}{\{2\pi npq\}^{\frac{1}{2}}}e^{-\delta_k^2/2npq}=\frac{1}{(npq)^{\frac{1}{2}}}\phi\left(\frac{\delta_k}{(npq)^{\frac{1}{2}}}\right). \quad (2.8)$$

这就是我們所要求的漸近公式．我們可以采用一个較方便的符号把它加以簡化．令

$$h=\frac{1}{(npq)^{\frac{1}{2}}}, \quad (2.9)$$

并且定义变量 k 的函数 x_k 如下：

$$x_k=(k-np)h=\frac{\delta_k}{(npq)^{\frac{1}{2}}}. \quad (2.10)$$

利用这些符号，我們可以把(2.8)写为

$$b(k;n,p)\sim h\phi(x_k). \quad (2.11)$$

在这个式子的推导中，我們假定了 $n\to\infty$ 和 $k\to\infty$ 满足 $\delta_k n^{-1}\to 0$ 和 $\delta_k^3 n^{-2}\to 0$．显然，最后一个条件蘊含了前面一个条件，而且它們等价于 $x_k^3 n^{-\frac{1}{2}}\to 0$．因此我們有

定理 1. 如果 $n\to\infty$ 和 $k\to\infty$ 以 $x_k^3 n^{-\frac{1}{2}}\to 0$ 的方式变化，則漸近公式(2.11)成立．更精确地說：我們証明了存在常数 A 和 B 使得

$$\left|\frac{b(k;n,p)}{h\phi(x_k)}-1\right|<\frac{A}{n}+\frac{B|x_k^3|}{n^{\frac{1}{2}}} \quad (2.12)$$

((2.11)的另一形式参看問題 19 和 21)．

图 3 对 $n=10$, $p=0.2$ 而 npq 只是 1.6 的情形来解释这个定理，可以看到即使在这样极端不利的情形下，逼近的程度还是很好[1]．

由我們的定理直接导出和数 (2.1) 的一个簡易的近似式．如果

$$hx_\alpha^3\to 0,\quad hx_\beta^3\to 0, \quad (2.13)$$

1) 当 $k=0,1,2,\cdots,6$ 时，对应的 $b(k;10,0.2)$ 的値分别为 0.1074, 0.2684, 0.3020, 0.2013, 0.0880, 0.0264, 0.0055；而对应的近似式 $h\phi(x_k)$ 的値分别为 0.0904, 0.2307, 0.3154, 0.2307, 0.0904, 0.0189, 0.0021.

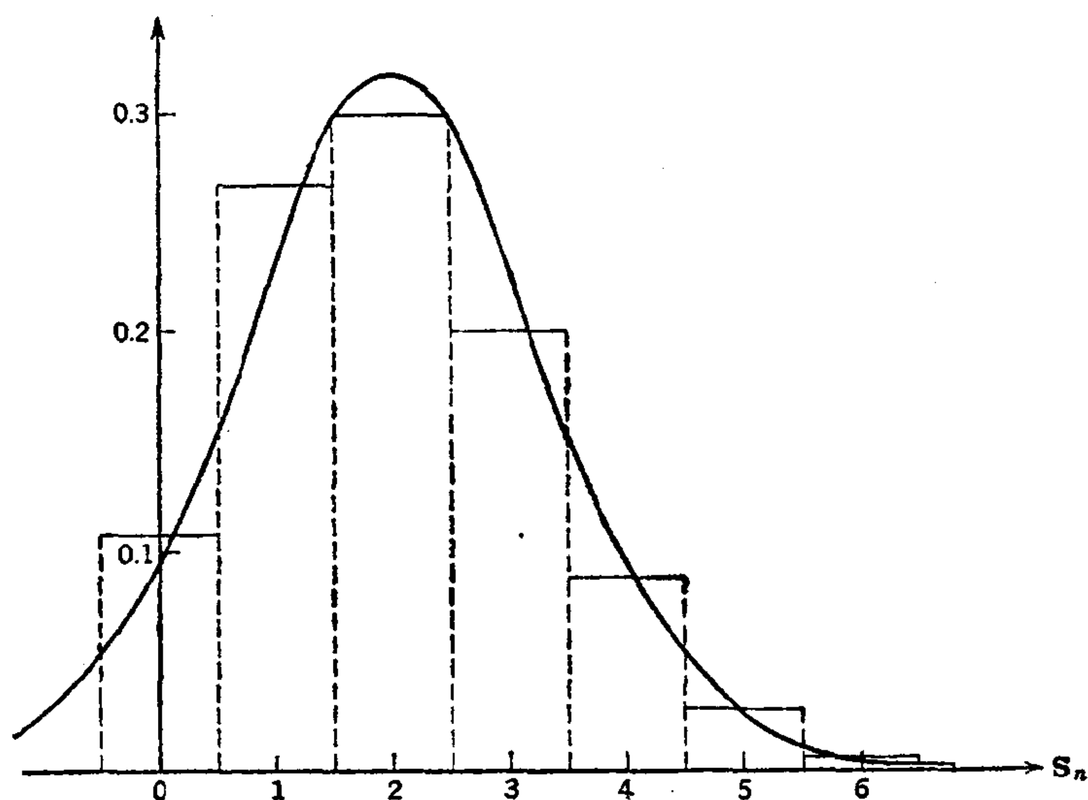


图 3. 伯努利分布的正态逼近，其中的阶梯函数表示具有 $p=\frac{1}{5}$ 的 10 次伯努利試驗中 k 次成功的概率 $b\left(k;10,\frac{1}{5}\right)$. 而那条連續曲綫则表示对于每个整数 k 所对应的正态逼近.

則对(2.1)中的所有各項,(2.11)都一致地成立. 因此

$$\mathbf{P}\{\alpha \leqslant \mathbf{S}_n \leqslant \beta\} \sim h\{\phi(x_\alpha)+\phi(x_{\alpha+1})+\cdots+\phi(x_\beta)\}. \tag{2.14}$$

右端为一个积分近似式的黎曼和[1],我們現在来研究其逼近的精确性.

由中值定理我們知道,存在一个 ξ_k 使

$$\Phi(x_{k+\frac{1}{2}})-\Phi(x_{k-\frac{1}{2}})=h\phi(\xi_k),$$

$$x_k-\frac{1}{2}h<\xi_k<x_k+\frac{1}{2}h. \tag{2.15}$$

因此

$$h\phi(x_k)=e^{\frac{1}{2}(\xi_k^2-x_k^2)}\{\Phi(x_{k+\frac{1}{2}})-\Phi(x_{k-\frac{1}{2}})\}. \tag{2.16}$$

---

1) 显然，$\Phi(x_{k+\frac{1}{2}})-\Phi(x_{k-\frac{1}{2}})$ 表示区間 $\left(x_k-\frac{1}{2}h, x_k+\frac{1}{2}h\right)$上曲綫$y=\phi(x)$下面的面积，$h\phi(x_k)$ 表示同一区間上对应的矩形的面积.

选取任意一个 $\varepsilon > 0$. 如果(2.13)成立，則对全部 $\alpha \leqslant k \leqslant \beta$ 和充分大的 n 都有

$$\frac{1}{2}|\xi_k^2 - x_k^2| = \frac{1}{2}|\xi_k - x_k| \cdot |\xi_k + x_k| < h\left[|x_k| + \frac{1}{4}h\right] < \varepsilon,$$

因此

$$e^{-\varepsilon}\{\Phi(x_{k+\frac{1}{2}}) - \Phi(x_{k-\frac{1}{2}})\} < h\phi(x_k) < \\ < e^{\varepsilon}\{\Phi(x_{k+\frac{1}{2}}) - \Phi(x_{k-\frac{1}{2}})\}. \tag{2.17}$$

对 k 求和，我們发現(2.14)的右边与 $\Phi(x_{\beta+\frac{1}{2}}) - \Phi(x_{\alpha-\frac{1}{2}})$ 之比趋于 0. 因此，我們証明了

德模佛-拉普拉斯极限定理. 如果 α, β 依照 $hx_\alpha^3 \to 0$, 及 $hx_\beta^3 \to 0$ 的方式变化，則

$$\mathbf{P}\{\alpha \leqslant \mathbf{S}_n \leqslant \beta\} \sim \Phi(x_{\beta+\frac{1}{2}}) - \Phi(x_{\alpha-\frac{1}{2}}), \tag{2.18}$$

其中 $h = (npq)^{-\frac{1}{2}}$, $x_t = (t - np)h$. 換言之，(2.18)两端之相对誤差随 hx_β^3, hx_α^3 的趋于零而趋于零.

特别，若限制 α 和 β 的值，使得对应的 x_α 与 x_β 落在某一固定的区間內，則(2.18)成立.（对 α 和 β 之取值相当大以致条件(2.13)不能滿足的情形，將在第 5 节及問題 14 中討論.）

在統計学中应用(2.18)时，通常是令 α 及 β 取这样的值，使得 $|x_\alpha|$ 及 $|x_\beta|$ 的值不超过 3 或 4. 然而在理論上应用(2.18)时，往往要使区間 (α, β) 远离二項分布的中心部分，这时对应的 x_α, x_β 都很大. 这时(2.18)的两边都很小. 因此去了解它們之比是否接近 1，与它們的差趋于零是同等的重要的.

如果我們引进"正則化成功次数"

$$\mathbf{S}_n^* = \frac{\mathbf{S}_n - np}{(npq)^{\frac{1}{2}}} \tag{2.19}$$

来代替 $\mathbf{S}_n$，則极限定理(2.18)会变得更加簡单. $\mathbf{S}_n^*$ 即为用 $(npq)^{\frac{1}{2}}$ 作单位去度量 $\mathbf{S}_n$ 与 np 之偏离. 数 np 称为 $\mathbf{S}_n$ 的均值，$(npq)^{\frac{1}{2}}$ 称为 $\mathbf{S}_n$ 的标准差. 用这述語是由随机变量的理論中所提示出来的（参看第九章）. 不等式 $\alpha \leqslant \mathbf{S}_n \leqslant \beta$ 与 $x_\alpha \leqslant \mathbf{S}_n^* \leqslant x_\beta$ 表示同样的事件，所以对任意固定的 $x_\alpha < x_\beta$，(2.18)可以表成

$$\mathbf{P}\{x_\alpha \leqslant \mathbf{S}_n^* \leqslant x_\beta\} \sim \Phi\left(x_\beta + \frac{h}{2}\right) - \Phi\left(x_\alpha - \frac{h}{2}\right), \quad (2.20)$$

其中 $h = (npq)^{\frac{1}{2}}$. 但当 $n \to \infty$ 时 $h \to 0$, 所以右边趋于 $\Phi(x_\beta) - \Phi(x_\alpha)$. 于是我們得到

极限定理的推論. 对每一組固定的 $a < b$, 都有

$$\mathbf{P}\{a \leqslant \mathbf{S}_n^* \leqslant b\} \to \Phi(b) - \Phi(a). \quad (2.21)$$

这是 (2.18) 的一个較弱的形式，然而它却是拉普拉斯极限定理的一个传統的形式. 如果略去 (2.20) 中的微量 $\frac{h}{2}$, 則带来的誤差随 $n \to \infty$ 而趋于 0. 但是当 npq 不太大的时候(如例(3.a)—(c)), 它却有着重大影响.

(2.21)所說明的主要事实是: 对于相当大的 n, 左边的概率在实际上与 p 无关. 这就可以使我們用标准的尺度去比較各組不同伯努利試驗序列的起伏.

定理 (2.21) 在历史上是概率的第一个极限定理. 从近代的观点来看, 它仅仅是中心极限定理的一个非常特殊的情形. 关于中心极限定理将在第十章中研究, 而其最一般的推导則要推迟到第二卷. 統計学家甚至对 npq 相当小时还用 (2.21) 来作逼近, 这时就需要估計誤差, 在許多情形下, (2.11) 的誤差比在 (2.14) 中和数用积分来代替时所产生的誤差要小(幸好, 可以用尤拉-麦克劳林 (Euler-MacLaurin) 求和公式避免后一誤差). 塞格·伯因斯坦发表了許多論文来研究在一般情况下的誤差, 并討論了如何修改 x_t 的定义以便使(2.18)的收斂性得到改善. 他的論文都是用俄文发表的, 而且不容易得到. 但是他的結果的改进的簡化推导, 已經有了英文文献[54].

关于任意停止的附注. 必須注意: 我們的极限定理和逼近定理只有当試驗次数 n 不依賴試驗的結果而事先固定时才是对的. 如果一个賭徒有权在对他有利的时刻停止賭博, 則他的最終贏得不能由正态逼近来判断, 因为这时賭博的持續时間是随机的. 对于每一个固定的 n 来說, $\mathbf{S}_n^*$ 很大的可能性很小. 然而, 在一个长的連貫中, 甚至不太可能的事件也可能发生, 而且我們将要看到: 在一場持續的賭博中, $\mathbf{S}_n^*$ 在实际上确有一序列极大值其阶为 $(\log\log n)^{\frac{1}{2}}$ (这就是第八章第 5 节的迭对数法則).

3. 例

(a) 令 $p=\frac{1}{2}$, $n=200$, $\alpha=95$, $\beta=105$, 在这种情形下, $\mathbf{P}\{95\leqslant \mathbf{S}_n\leqslant 105\}$ 可以解释为把一个錢币扔 200 次出現正面的次数与 100 之偏差不超过 5 的概率. 我們有 $h=(50)^{-1/2}=0.141421\cdots$ 和 $-x_{\alpha-\frac{1}{2}}=x_{\beta+\frac{1}{2}}=(5.5)h=0.7778\cdots$. 查表即得 $\Phi(x_{\beta+\frac{1}{2}})-\Phi(x_{\alpha-\frac{1}{2}})=0.56331\cdots$. 眞正的值(也是由查表而得的)是 $0.56325\cdots$. 这个誤差小得惊人, 这是因为在我們这个問題所考虑的区間中, 积分是(2.14)中和数的高估計, 而(2.11)又是对每一項的低估計.

(b) 令 $p=\frac{1}{10}$, $n=500$, $\alpha=50$, $\beta=55$. 正确的值 $\mathbf{P}\{50\leqslant \mathbf{S}_n\leqslant 55\}=0.317573\cdots$. 但是 $h=(45)^{-\frac{1}{2}}=0.1490712\cdots$, 可得出近似值 $\Phi(5.5h)-\Phi(-0.5h)=0.3235\cdots$. 其誤差大約为百分之二.

表 2. 对 $n=100$, $p=0.3$ 的二項分布与正态逼近的比較表

成功次数	概　率	正态逼近	相对誤差
$9\leqslant \mathbf{S}_n\leqslant 11$	0.000 006	0.000 03	+400
$12\leqslant \mathbf{S}_n\leqslant 14$	0.000 15	0.000 33	+100
$15\leqslant \mathbf{S}_n\leqslant 17$	0.002 01	0.002 83	+40
$18\leqslant \mathbf{S}_n\leqslant 20$	0.014 30	0.015 99	+12
$21\leqslant \mathbf{S}_n\leqslant 23$	0.059 07	0.058 95	0
$24\leqslant \mathbf{S}_n\leqslant 26$	0.148 87	0.144 47	−3
$27\leqslant \mathbf{S}_n\leqslant 29$	0.237 94	0.234 05	−2
$31\leqslant \mathbf{S}_n\leqslant 33$	0.230 13	0.234 05	+2
$34\leqslant \mathbf{S}_n\leqslant 36$	0.140 86	0.144 47	+3
$37\leqslant \mathbf{S}_n\leqslant 39$	0.058 89	0.058 95	0
$40\leqslant \mathbf{S}_n\leqslant 42$	0.017 02	0.015 99	−6
$43\leqslant \mathbf{S}_n\leqslant 45$	0.003 43	0.002 83	−18
$46\leqslant \mathbf{S}_n\leqslant 48$	0.000 49	0.000 33	−33
$49\leqslant \mathbf{S}_n\leqslant 51$	0.000 05	0.000 03	−40

(c) 令 $n = 100$, $p = 0.3$. 表 2 以一个典型的例子（对相对地較小的 n）說明：当 (α, β) 远离中心項时正态逼近怎样地在变坏.

(d) 計我們找寻这样的一个数 a, 使得对于相当大的 n 由不等式 $|\mathbf{S}_n^*| > a$ 所确定的事件具有接近于 $\frac{1}{2}$ 的概率，为此必需 $\Phi(a) - \Phi(-a) = \frac{1}{2}$, 或者 $\Phi(a) = \frac{3}{4}$. 从正态分布表中可以找到 $a = 0.6745$，因此两个不等式

$$
\begin{aligned}
&|\mathbf{S}_n - np| < 0.6745(npq)^{\frac{1}{2}}, \\
&|\mathbf{S}_n - np| > 0.6745(npq)^{\frac{1}{2}}
\end{aligned}
\tag{3.1}
$$

的可能性大体上是相等的. 特别，把一个錢币扔 n 次出現"正面"的次数落在区間 $\left[\frac{n}{2} - 0.337n^{\frac{1}{2}}, \frac{n}{2} + 0.337n^{\frac{1}{2}}\right]$ 內的概率大約为 1/2. 同理，把一顆骰子擲 n 次，出現"么点"的次数落在区間 $\left[\frac{n}{6} - 0.251n^{\frac{1}{2}}, \frac{n}{6} + 0.251n^{\frac{1}{2}}\right]$ 內的概率大約也是 $\frac{1}{2}$. $\mathbf{S}_n$ 落在区間 $[np - 2(npq)^{\frac{1}{2}}, np + 2(npq)^{\frac{1}{2}}]$ 內之概率約为 $\Phi(2) - \Phi(-2) = 0.9545\cdots$; 若把区間改为 $[np - 3(npq)^{\frac{1}{2}}, np + 3(npq)^{\frac{1}{2}}]$，則对应的概率为 $0.9973\cdots$.

(e) 一个竞争問題. 这是說明公式 (2.21) 的实际应用的一个例子. 在芝加哥与洛杉磯之間有两条竞爭的鉄路，它們的火車是同时开出同时到达并且具有同样的設备. 假定 n 个旅客乘坐哪一条鉄路的火車的选择是互相独立的而且又是任意的，于是每列火車的乘客数目可視为以概率 $p = \frac{1}{2}$ 的 n 次伯努利試驗的結果. 如果一列火車設置 $s < n$ 个座位、那么，一旦有多于 s 个旅客要来乘車，就容納不下了. 令这个事件的发生概率为 $f(s)$，且 $f(s) > 0$，应用漸近公式 (2.21) 就有

$$
f(s) \approx 1 - \Phi\left(\frac{2s - n}{n^{\frac{1}{2}}}\right). \tag{3.2}
$$

如果 s 是如此的大，以致 $f(s) < 0.01$，那么在 100 次中有 99 次是

有足够的座位的. 一般說来，公司可以任意規定一个冒险水准 α 并这样来决定 s 使 $f(s) < \alpha$. 为此，只要令

$$s \geqslant \frac{1}{2}(n + t_\alpha n^{\frac{1}{2}}), \qquad (3.3)$$

其中 t_α 是方程 $\alpha = 1 - \Phi(t_\alpha)$ 的根，可在表中查得. 例如若 $n = 1000$, $\alpha = 0.01$, 則 $t_\alpha \approx 2.33$, 而 $s = 537$ 个座位就足够了. 如果两条鉄路都規定冒险水准为 $\alpha = 0.01$, 則两列火車所有的座位的总数为 1074，其中有 74 个空座位. 可見由于竞爭(或者机会的波动)而带来的損失是很小的. 用同样的方法可算出，若有 514 个座位則在 80% 的情况下是够的；而若有 549 个座位，則在 999/1000 的情况下是够的.

类似的考虑可以应用到其它的有竞爭的供应問題上去，例如对 n 位顾客有 m 个电影院来竞爭，于是每个电影院成功的概率是 $p = \frac{1}{m}$，而(3.3)应該用 $s \geqslant \left(\frac{1}{m}\right)[n + t_\alpha n^{\frac{1}{2}}(m - 1)^{\frac{1}{2}}]$ 来代替. 这时空座位的总数約为 $ms - n \approx t_\alpha n^{\frac{1}{2}}(m - 1)^{\frac{1}{2}}$. 当 $\alpha = 0.01$, $n = 1000$, $m = 2, 3, 4$ 时，空座位的数目大約分别为 74，126，147. 由于竞爭而带来的使用率的損失是非常小的.

(f) 随机数字. 在第二章例(3.b)中，曾經考虑具有 $p = 0.3024$ 的事件，在 $n = 1200$ 次試驗中，这个事件的平均頻率为 0.3142，它与 p 的离差为 $\varepsilon = 0.0118$. 在这里，$(pq)^{\frac{1}{2}} = 0.4593$, $\varepsilon(n/pq)^{\frac{1}{2}} \approx 0.890\cdots$. 因此 $\left|\frac{S_n}{n} - p\right| > \varepsilon$ 的概率大約为$0.37\cdots$，它說明大約占总情况的 37% 成功的平均数与 p 之偏差大于 ε.

(g) 抽样. 在某一羣居民中，吸烟的人的比数为 p. 假定 p 是未知数而用有放回的随机抽样来决定 p，要求其誤差不超过 0.005. 为此样本的大小 n 究竟要多大呢？ 如果 p' 为(抽出来的)样本中吸烟人的比数，我們希望 $|p' - p| < 0.005$. 但是不論样本有多大，总不能絕对地保証：$|p' - p| < 0.005$；样本中人人吸烟也是可以設想的，因为絕对的必然是不可能达到的. 我們取一个"置信

水准”α，比方說：$\alpha = 0.95$，即要求 $|p' - p| < 0.005$ 的概率大于等于 0.95. 注意：np' 是 n 次試驗成功的总数，因此

$$\mathbf{P}\{|p' - p| < 0.005\} = \mathbf{P}\left\{\left|\frac{\mathbf{S}_n}{n} - p\right| < 0.005\right\}.$$

我們求 n 要足够大，使得上列的概率大于 0.95. 为此，只要利用正态逼近就成了. 方程 $\Phi(x) - \Phi(-x) = 0.95$ 的根 x 等于 1.96 $\cdots$，因而有 $0.005\left(\frac{n}{pq}\right)^{\frac{1}{2}} \geqslant 1.96\cdots$，即 $n \geqslant (392)^2 pq$，或近似地 $n \geqslant 160{,}000pq$. 这个不等式虽然依赖于未知量 p，而 pq 决不超过 $\frac{1}{4}$，因此样本的大小取作 $n = 40{,}000$ 时，則在一切情形下都可以了；而且 $|p' - p| < 0.005$ 成功的机会 19 倍于失敗.

4. 与普阿松逼近的关系

当 npq 相当大时，正态逼近的誤差很小. 另一方面，当 n 大而 p 小时，则 $b(k; n, p)$ 将接近于 $\lambda = np$ 的普阿松概率 $p(k;\lambda)$. 如果 λ 很小，那就只能用普阿松逼近了. 然而当 λ 較大时，則正态逼近与普阿松逼近都可以应用. 由此可知，对 λ 較大的値，就可能用正态分布去逼近普阿松分布，且在第十章例 (1.c) 中我們将看到确实如此（也可参看問題 9）. 但在这里，我們只用一个数値的例子和一个实际的例子来說明这一点.

例. (a) 考虑普阿松分布 $p(k;100)$. 我們可以用它来逼近一个具有 $n = 100{,}000{,}000$，$p = 1/1{,}000{,}000$ 的二項分布. 此时 $npq \approx 100$. 虽然这个量并不大，但至少在二項分布的中心部分可以用正态分布来很好地逼近它. 普阿松分布 $p(k;100)$ 与 $b(k; 10^8, 10^{-6})$ 的前面好几位小数是相同的，可以把它和后者的近似正态分布来作比較. 为簡单起見，令 $P(a, b) = p(a; 100) + p(a + 1; 100) + \cdots + p(b; 100)$，于是，$P(a,b)$ 替代了 $\mathbf{P}\{a \leqslant \mathbf{S}_n \leqslant b\}$，而且可用

$$\Phi\left(\frac{b - 99.5}{10}\right) - \Phi\left(\frac{a - 100.5}{10}\right)$$

来近似．下面几个数字給出逼近程度的輪廓．

	正确的值	正态逼近
$P(85,90)$	0.113 84	0.110 49
$P(90,95)$	0.184 85	0.179 50
$P(95,105)$	0.417 63	0.417 68
$P(90,110)$	0.706 52	0.706 28
$P(110,115)$	0.107 38	0.110 49
$P(115,120)$	0.053 23	0.053 35

(b) 电話綫路問題． 下面的問題是从实践中經过一些簡化以后提出来的[1]．一个电話交換台 A 为靠近于交換台 B 的 2000 个用戶服务．如果从 A 到 B 安装 2000 条綫路，那就太貴太浪費了．我們只要安装的綫路的数目 N 是如此大，使得在通常情形下，100 次呼喚至多只有 1 次不能立刻接通．假定在一天最忙的时間里每个用戶打到 B 区去的电話在一点鈡內平均占綫两分鈡．不妨把最忙碌的时間里的一个固定时刻的情况比拟为 $p=\frac{1}{30}$ 的 2000 次試驗，其中 $p=\frac{1}{30}$ 表示每戶用綫的概率．在通常的情况下，这些試驗可視为相互独立的．（这并不总是正确的． 如果由于意外的大雨或地震，很多人都要叫出租汽車或本地报館的电話，綫路就会发生“拥挤”，那时，独立性的理論就不能用了．）因而就得到了具有概率 $p=\frac{1}{30}$ 的 2000 次伯努利試驗，求满足下面的条件的最小的設置綫路数目 N，使得“成功”的次数（卽占綫的数目）超过 N 的概率小于 0.01，用記号表示： $\mathbf{P}\{\mathbf{S}_{2000}\geqslant N\}<0.01$．

对普阿松逼近来說，取 $\lambda=200/3\approx66.67$． 查表可得，成功次数大于等于 87 的概率約为 0.0097，然而成功次数大于等于 86 的概率約为 0.013．由此可見安装 87 条綫路就足够了．对正态逼近来說我們首先要找出方程 $1-\Phi(x)=0.01$ 的根 x，由查表可

1) 参看 [55]．在該书中此問題是用普阿松方法来处理的，工程师总是喜欢用普阿松方法．

得 $x = 2.327$. 于是我們要求: $\left(N-\frac{1}{2}-np\right)\Big/(npq)^{\frac{1}{2}}\geqslant 2.327$.
由于 $n = 2000$, $p = 1/30$, 所以 $N \geqslant 67.17 + (2.327)(8.027) \approx 85.8$. 因此,使用正态逼近法的結果是: 只需安装 86 条綫路就足够了.

这两个解答可以說是一致的. 这种方法还可以給出另外一些有用处的結果. 可以这样設想: 把 2000 个用戶分成两組,每組各 1000 戶, 从 A 到 B 的綫路也与之对应地分成两組来安装. 用上述方法算一算結果还要多置 10 条綫路,因此第一种安装法还是更好一些.

5. 大 偏 差[1]

我們常常要估計"正則化的成功次数 $\mathbf{S}_n^*$"(参看(2.19))大于某个給定的数值 x 的概率. 这时区間的上端是无穷大, 因此需要做一些特殊的論証来說明我們的极限定理(2.18)仍旧适用.

定理. 当 $n\to\infty$ 时若 x 作为 n 的一个函数而变化, 使得 $x\to\infty$, $x^3h\to 0$ 則

$$\mathbf{P}\{\mathbf{S}_n^* > x\}\sim 1-\Phi(x). \tag{5.1}$$

由(1.7), (5.1)等价于

$$\mathbf{P}\{\mathbf{S}_n^* > x\}\sim \frac{1}{(2\pi)^{\frac{1}{2}}x}e^{-\frac{1}{2}x^2}. \tag{5.2}$$

証. 在(2.18)中选取整数 α, β,使 x 落在 x_α 与 $x_{\alpha+1}$ 之間,同时 $x_\beta\approx x+\log x$. 于是 $x_\beta^3h\to 0$ 从而(2.18)成立. 因此

$$\mathbf{P}\{\alpha<\mathbf{S}_n<\beta\}\sim\{1-\Phi(x_\alpha)\}-\{1-\Phi(x_\beta)\}. \tag{5.3}$$

然而从(1.7)及 $x_\beta\approx x_\alpha+\log x_\alpha$ 容易看出: $1-\Phi(x_\beta)$ 的数量級比 $1-\Phi(x_\alpha)$ 要小,同时 $1-\Phi(x_\alpha)\sim 1-\Phi(x)$. 所以

$$\mathbf{P}\{\alpha<\mathbf{S}_n<\beta\}\sim 1-\Phi(x). \tag{5.4}$$

---

1) 本节中的定理具有多方面的意义, 然而在本书中只有在第八章第 5 节中关于迭对数法則的証明中才用到它.

另一方面，从(2.11)及第六章公式(3.5)我們有

$$\mathbf{P}\{\mathbf{S}_n \geqslant \beta\} \leqslant \frac{n}{\beta - np} b(\beta; n, p) \sim \frac{nh^2}{x_\beta} \phi(x_\beta). \tag{5.5}$$

因为 $nh^2 = 1/pq$ 是一个常数，并且

$$\frac{1}{x_\beta} \phi(x_\beta) \sim 1 - \Phi(x_\beta). \tag{5.6}$$

我們已經看到，右边趋于零的速度較 $1 - \Phi(x)$ 为快，这就意味着 $\mathbf{P}\{\mathbf{S}_n \geqslant \beta\}$ 的数量級要比 $1 - \Phi(x)$ 还要小．把这些結果与(5.4)結合起来，則得

$$\mathbf{P}\{\mathbf{S}_n > \alpha\} \sim 1 - \Phi(\alpha), \tag{5.7}$$

这就是我們要想証明的定理．(关于大偏差的另外一些极限定理将在問題 12—17 中給出．)

6. 問　　題

1．推广(1.8)，証明：

$$1 - \Phi(x) \sim \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}x^2} \left\{\frac{1}{x} - \frac{1}{x^3} + \frac{1\cdot 3}{x^5} - \frac{1\cdot 3\cdot 5}{x^7} + - \cdots + \right.$$

$$\left. + (-1)^k \frac{1\cdot 3 \cdots (2k-1)}{x^{2k+1}}\right\}, \tag{6.1}$$

且当 $x > 0$ 时，若 k 为偶数則(6.1)之右端高估了 $1 - \Phi(x)$；若 k 为奇数則(6.1)右端低估了 $1 - \Phi(x)$．

2．对于每个常数 $a > 0$，当 $x \to \infty$ 时

$$\left\{1 - \Phi\left(x + \frac{a}{x}\right)\right\} \div \{1 - \Phi(x)\} \to e^{-a}. \tag{6.2}$$

3．求出 10,000 个随机数字中，数字 7 出現的次数不多于 968 次的概率．

4．把一顆骰子，連續擲 12,000 次，求出現“么点”的次数在 1900 与 2150 之間的概率的近似值．

5．求一个数 k，使得扔 1000 次錢币出現“正面”的次数在 440 与 k 之間的概率約为 0.5．

6．为了求出女性居民在居民中所占的比数 f，进行抽样調査，問样本要取多大方能使抽样誤差小于 0.005 的概率大于等于 0.99?

7．一个錢币扔了 10,000 次，出現了 5400 次“正面”，我們是否能够断言

这个錢币不均匀呢?

8. 求三項分布

$$\frac{n!}{k!\,r!\,(n-k-r)!}\,p_1^k p_2^r (1-p_1-p_2)^{n-k-r}$$

中最大項的近似值。

9. 普阿松分布的正态逼近。利用史特令公式証明: 如果 $\lambda\to\infty$, 則对于任意固定的 $\alpha<\beta$

$$\sum_{\lambda+\alpha\lambda^{\frac{1}{2}}<k<\lambda+\beta\lambda^{\frac{1}{2}}} p(k;\lambda)\to\Phi(\beta)-\Phi(\alpha). \tag{6.3}$$

10. 超几何分布的正态逼近。 令 m, n, k 为正整数且假定

$$\frac{r}{n+m}\to t,\quad \frac{n}{n+m}\to p,\quad \frac{m}{n+m}\to q,\quad h\{k-rp\}\to x \tag{6.4}$$

其中 $1/h=(n+m)pqt(1-t)^{\frac{1}{2}}$. 試証:

$$\frac{\binom{n}{k}\binom{m}{r-k}}{\binom{n+m}{r}}\sim h\phi(x). \tag{6.5}$$

提示: 用二項分布的正态逼近比用史特令公式好。

11. 正态分布及組合連貫[56]。 在第二章(11.19)中我們看到, 在 n 个 α 与 m 个 β 的排列中, 恰有 k 个 α 連貫的概率为

$$\pi_k=\binom{n-1}{k-1}\binom{m+1}{k}\div\binom{n+m}{n}. \tag{6.6}$$

令 $n\to\infty$, $m\to\infty$, 并設(6.4)成立. 对于固定的 $\alpha<\beta$, α 連貫的个数落在 $npq+\alpha(npq)^{\frac{1}{2}}$ 与 $npq+\beta(npq)^{\frac{1}{2}}$ 之間的概率趋于 $\Phi(\beta)-\Phi(\alpha)$.

在下面的問題中, $h^2=npq$, $\mathbf{S}_n^*$ 是(2.19)所定义的"正則化的成功次数". 而

$$F_n(x)=\mathbf{P}\{\mathbf{S}_n^*>x\}. \tag{6.7}$$

12. 若 x 作为 n 的函数而变化, 使得

$$x^{3+a}h\to 0,\qquad x\to\infty,$$

則[57]

$$\frac{F_n(x)}{1-\Phi(x)}=1+o(x^a), \tag{6.8}$$

其中 $o(x^a)$ 代表数量級比 x^a 小的量。

13. 若 $x^3h\to 0$, $x\to\infty$. 則[58]对于任意常数 $a>0$, 有

$$\frac{F_n(x) - F_n\left(x + \frac{a}{x}\right)}{F_n(x)} \to 1 - e^{-a}. \tag{6.9}$$

卽是說，在条件 $S_n^* > x$ 的条件下 $x < S_n^* < x + \frac{a}{x}$ 的条件概率趋向于 $1 - e^{-a}$ [提示：应用(5.2)].

14. 大偏差的概率. 从(2.4)出发証明下面的定理. 如果 $n \to \infty$，且变化 k 使得 $(k - np)/n \to 0$，則

$$b(k; n, p) \sim \frac{h}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}x^2 - f(x)}. \tag{6.10}$$

其中 $x = (k - np)h$，且

$$f(x) = \sum_{\nu=3}^{\infty} \frac{p^{\nu-1} - (-q)^{\nu-1}}{\nu(\nu - 1)} h^{\nu-2} x^{\nu}. \tag{6.11}$$

注意：如 $x^3 h \to 0$，則 $f(x) \to 0$，从而(6.10)可化成(2.11). 如果 x 的数量級为 $h^{-\frac{1}{3}}$，但 x 与 $h^{-\frac{1}{2}}$ 比起来却是可忽略的，則有

$$f(x) \approx \frac{p - q}{6} x^3 h. \tag{6.12}$$

如果 x 的数量級为 $h^{-\frac{1}{2}}$，則有

$$f(x) \approx \frac{p - q}{6} x^3 h + \frac{p^3 + q^3}{12} x^4 h^2, \tag{6.13}$$

等等.

15(續前). 証明：如果 $x \to \infty$，$xh \to 0$，則

$$f\left(x + \frac{a}{x}\right) - f(x) \to 0, \tag{6.14}$$

且有

$$F_n(x) \sim e^{f(x)} \{1 - \Phi(x)\}. \tag{6.15}$$

16. 在只假定 $xh \to 0$ 的条件下，由(6.15)推出(6.9).

17. 若 $p > q$，則对于相当大的 x

$$\mathbf{P}\{\mathbf{S}_n < x\} < \mathbf{P}\{\mathbf{S}_n < -x\} \tag{6.16}$$

(提示：应用問題 14).

18. 大数定律的一个新推导. 試証大数定律可由德模佛-拉普拉斯极限定理推出.

19.正态逼近的一个新推导[1]. 从第六章 (10.11)—(10.13) 出发，証明：当 $n\to\infty$ 和 $k\to\infty$ 满足 $\xi_k^3 n^{-\frac{1}{2}}\to 0$ 时,我們有

$$b(k;\,n,\,p)\sim b(m;\,n,\,p)e^{-\frac{1}{2}\xi_k^2}. \tag{6.17}$$

20.如果 $np\leqslant m\leqslant (n+1)p$, 証明:

$$b\left(m;\,n,\,\frac{m+1}{n+1}\right)\leqslant b(m;\,n,\,p)\leqslant b\left(m;\,n,\,\frac{m}{n}\right). \tag{6.18}$$

如果 $(n+1)p-1<m\leqslant np$, 同样的不等式仍然成立，只不过把最左边的 $\frac{m+1}{n+1}$ 代之以 $m/(n+1)$ 罢了.

21.推出 $b(m;n,p)\sim\{2\pi(n+1)pq\}^{-\frac{1}{2}}$,并給出上界和下界.

---

1) 問題 19 和 20 加在一起蘊含了 $b(k;\,n,\,p)\sim\{(n+1)pq\}^{-\frac{1}{2}}\phi(\xi_k)$. 这和基本的逼近公式 (2.11) 是一样的，只不过以 ξ_k 代 x_k, 以 $h'=\{(n+1)pq\}^{-\frac{1}{2}}$ 代 $h=\{npq\}^{-\frac{1}{2}}$. 因为 $x_k\sim\xi_k$ 和 $h\sim h'$,所以两个公式是漸近等价的. 实际上,新公式的誤差項比較小. (在它的推导中,由(2.7)到 (2.8) 的誤差可以避免.) 还須注意的是: 問題 19 和 20 所需要的計算比課文中所需要的更簡单更直观; 它們只包含如第二章第 8 节和第六章第 3 节所用过的关于对数的标准估計. 簡短地說，新公式及其推导都比課文中对应的公式及推导較为优越，只是它們不符合于长期以来采用的 np 而代以 $(n+1)p$.

*第八章　伯努利試驗的无穷序列

本章所討論的是关于伯努利試驗的随机性的某些性質，及关于它的重要的迭对数法則．伯努利試驗的起伏理論$\left(\text{至少对}p=\frac{1}{2}\right)$的一个不同的方面已包含在第三章中．

1. 試驗的无穷序列

在前一章中我們已經詳細地討論了关于 n 次伯努利試驗的概率，同时也研究了当 $n\to\infty$ 时，它們的漸近性質．現在我們轉而研究一类更为一般的問題，其中事件本身在有限的样本空間中幷不能定义．

例． 一个連貫問題．令 α, β 为两个正整数，我們考虑一个潛在的伯努利試驗的无穷序列，譬如扔一个錢币或者擲一顆骰子．設想甲与乙打这样一个賭：如果接連成功 α 次的連貫发生在接連失敗 β 次的連貫之前，就算甲贏．“甲贏”这个事件的概率，它的直观意义是很清楚的，但是要記住，在数学理論中事件这个术語是代表“样本点的集合”的，它只有在适当地定义了样本空間之后才有意义．为此，伯努利試驗的有限次（設 n 次）模型对我們現在的情形来說就不够了．但这个困难只需用簡单的取极限的办法就可以解决．在 n 次試驗以后，甲或則获胜或則失敗，或者不分胜負．令对应的概率分別为 x_n, y_n, z_n $(x_n+y_n+z_n=1)$．当試驗的次数 n 增加时，不分胜負的概率 z_n 只能下降，而 x_n 及 y_n 一定上升．因此 $x=\lim\limits_{n\to\infty}x_n$，$y=\lim\limits_{n\to\infty}y_n$，$z=\lim\limits_{n\to\infty}z_n$ 一定存在．不会有人迟疑而把 x, y, z 称为甲最后获胜、失敗或胜負不分的概率．然而，

---

* 标有星号的这一章与以后各章題材无直接联系，初讀时可以略去．

对应的三个事件只能定义在試驗的无穷序列所成的样本空間中，而且这个空間并不是离散的.

这个例子的引进仅仅为了說明問題，至于 x_n, y_n, z_n 的数值究竟是多少不是我們当前所关心的. 在第十三章例(8.b)中可以計算它們. 然而有一个比較簡单方法可以算出极限 x, y, z，这个方法在更一般的場合也可以应用. 由于这个方法重要，而且很有意思，我們現在就来講一講.

令 A 表示"接連成功 α 次的連貫的发生在接連失敗 β 次的連貫以前"这一事件. 于是 A 意味着甲获胜，而 $x=\mathbf{P}\{A\}$. 令 u, v 分别表示在"第一次試驗之結果为成功"或"失敗"的条件下事件 A 的条件概率. 于是 $x=pu+qv$ [第五章(1.8)]. 首先設第一次試驗的結果为成功. 于是事件 A 可以有下面 α 种互不相容的方式发生，分为二类来描述：(1)紧接在后面的 $\alpha-1$ 次試驗的結果都是成功，其对应的概率为 $p^{\alpha-1}$. (2)第一次失敗发生在第 ν 次試驗，其中 $2\leqslant\nu<\alpha$. 令这一事件为 H_ν. 于是 $\mathbf{P}\{H_\nu\}=p^{\nu-2}q$；而 $\mathbf{P}\{A|H_\nu\}=v$. 因此(再度应用复合概率公式)

$$u=p^{\alpha-1}+qv(1+p+\cdots+p^{\alpha-2})=p^{\alpha-1}+v(1-p^{\alpha-1}). \quad (1.1)$$

在第一次試驗的結果为失敗的情形下，用类似的討論可以得到

$$v=pu(1+q+\cdots+q^{\beta-2})=u(1-q^{\beta-1}). \quad (1.2)$$

这样一来，我們得到了关于未知数 u, v 的两个綫性方程，解出 u, v 后再用 $x=pu+qv$，即得

$$x=p^{\alpha-1}\frac{1-q^{\beta}}{p^{\alpha-1}+q^{\beta-1}-p^{\alpha-1}q^{\beta-1}}. \quad (1.3)$$

至于求 y 之計算，只須把(1.3)中之 p 与 q；α 与 β 之位置相互对换即可，于是

$$y=q^{\beta-1}\frac{1-p^{\alpha}}{p^{\alpha-1}+q^{\beta-1}-p^{\alpha-1}q^{\beta-1}}. \quad (1.4)$$

由于 $x+y=1$，所以 $z=0$，即是說胜負不分的概率为 0.

例如：扔一个錢币$\left(p=\frac{1}{2}\right)$，出現两次"正面"的連貫在出現三次"反面"的連貫之前的概率是 0.7；接連出現两次"正面"在接連出現四次"反面"之前的概率是 5/6；接連出現三次"正面"在接連出現四次"反面"之前的概率为 15/22。在擲骰子时，接連出現二次么点在接連出現五次非么点之前的概率是 0.1753，等等.

在这一卷书里我們只限于討論离散的样本空間的理論，这就大大地損害了数学理論的完美性. 在一般理論中，把 n 次伯努利

試驗仅仅作为一个无穷試驗序列的起头的一部分来考虑. 因而一个样本点是由字母 S（成功）和 F（失败）的一个无穷序列来代表，而样本空間就是所有这样的序列的总体. 一个有限序列譬如說 $SSFS$，它表示最初四次試驗分别为 S, S, F, S 的那些样本点的总体. 在无穷的样本空間里，上面关于賭博的那个例子不需要用极限过程就能說明了. 任取一点，也就是一个序列 $SSFSFF\cdots$. 在它里面，由接連 α 个 S 所构成的連貫可以出現，也可以不出現，如果出現的話，那末在这以前，也可以出現过"接連 β 个 F 所构成的連貫"，也可以沒有. 这样一来，所有的样本点可以分为三类分别表示事件："甲获胜"，"甲失败"，"不分胜負"，它們的概率就是上面算出来的 x, y, z. 在这个样本空間中唯一令人苦恼的是：它不是离散的样本空間，而我們对一般样本空間的概率又沒有定义.

請注意，我們所討論的其实并不是眞困难，而是一个术語的問題. 拿我們的例子来說，問題不在于数 x 的定义下得适当不适当，及对它的解释給的正确不正确，問題仅仅在于，为了避免自相矛盾，我們要么把 x 說成是"在 n 次試驗中甲贏的概率 x_n 的极限"，要么把它說成是"甲贏"的概率，而样本空間不是离散的. 我們打算两种办法都用. 为了語言的簡短，我們使用事件这一术語，即使它們是定义在无穷的样本空間中；为了理論的精确，叙述定理时将要先用有穷的样本空間，然后过渡到极限. 这一章里所研究的事件都具有上述例子的如下的特色. "甲获胜"这一事件虽然是定义在无穷的样本空間中的，但它可表为"甲在第 n 次試驗获胜"($n = 1, 2, \cdots$）的那些事件的并，它們中的每一个只依賴于有限次試驗. 我們所要求的 x 是单調序列 x_n 的极限，这些概率 x_n 只依賴于有限次試驗. 我們并不需要比 n 次伯努利試驗模型的理論更高的理論；我們只不过簡化了累贅的表达[1]，把本来是"概率的极限"的一

1) 如果讀者对一般測度論較熟悉的話，那么这个情形可以描述如下. 我們仅仅考虑这样一些事件；或者它只依賴于有限次試驗，或者它只是这类事件的单調序列的极限. 我們計算这些明显的概率的极限，显然这并不需要測度論. 然而只有在測度論里才能証明：这样的极限值不依賴于特殊的取极限的过程而且是完全可加的.

些数目叫作概率.

2. 賭博的长策

很多賭徒的慘痛經历使我們得到这样一个教訓: 沒有一种賭博长策能成功地改善賭徒的机遇. 如果概率論正确反映生活的話,这个經驗就应該对应于一个可以証明的命題.

为了确定起見, 让我們考虑一个潛在的伯努利試驗的无穷序列, 而且假定在每一次試驗中賭徒都有选择賭或者不賭的自由. 所謂賭博的"长策",就是賭徒事先制定的一套規則,来决定哪一次該賭. 例如,賭徒可以决定在每个第七次上賭, 或者在賭一次之后,再出現七次"正面"之后的那一次賭. 他也可以只在"正面"連續出現13次以后才賭;或者在第一次"正面"出現以后賭第一次,然后在头一回連續出現两次"正面"以后賭第二次,一般地,在連續出現 k 次"正面"以后就賭第 k 次. 在最后一种情况, 他賭的頻次将是越来越少的. 我們并不需要考虑个别試驗的賭注; 我們将要証明沒有一个"长策"可以改变賭徒的处境, 他們所得到的結果和每次都賭是一样的. 不言而喻,只能在通常意义下的"长策",即在賭徒不知未来的情况下才能証明这个命題. (至于眞正的未卜先知存在与否的問題并不是我們所关心的.)同时,象"輸了三次以后便回家"这类規則是可能改变賭徒的处境的,不过这些沒意思的长策我們把它除外.

长策的定义是一套規則, 按照这套規則, 賭徒在无論哪一次上,唯一地决定他是参加賭还是不参加賭. 在第 k 次試驗上,他的决策可以依賴于以前的 $k-1$ 次試驗的結果,但是不可以依賴于第 $k, k+1, k+2, \cdots$ 次試驗結果; 最后,这些規則还要求能够保証: 賭博可以无穷的継續进行下去. 因为規則是固定的, 所以"在 n 次試驗中賭徒参加賭的次数超过 r 次"的事件是有定义的, 而它的概率是可以計算出来的. 上面定义里的最后的条件要求对任何一个 r, 当 $n \to \infty$ 时这个概率趋于 1.

我們的基本定理是: 无論长策是什么, 賭徒参加賭博的那些

次試驗組成一个伯努利試驗序列其成功概率不变. 如果在措詞方面做适当的修改,这个定理对一切类型的独立試驗都成立;在任何情况下，进行賭博的那些次試驗都是原来試驗的一模一样的仿制品,所以沒有一个长策可以改变賭徒的运气. 这个命題的重要性首先由馮·米賽斯所发現、他把它作为基本公理而引进，即所謂有效的賭博长策的不可能性. 这里的陈述及証明都是引用杜勃[59]的. 为了簡单起見,我們假定 $p=\frac{1}{2}$.

令 A_k 表示“第一次賭博发生在第 k 次試驗”这一事件. 长策的定义要求：当 $n\to\infty$ 时，在第 n 次試驗以前进行第一次賭博的概率趋于 1. 也就是 $\mathbf{P}\{A_1\}+\mathbf{P}\{A_2\}+\cdots+\mathbf{P}\{A_n\}\to 1$, 或

$$\sum\mathbf{P}\{A_k\}=1. \tag{2.1}$$

其次,令 B_k 表示“第 k 次试验出正面”的事件. 于是“第一次参加賭的那次試驗里出現正面”的事件 B 是事件 $A_1B_1, A_2B_2, A_3B_3, \cdots$ 的并，而这些事件是互斥的. 因为 A_k 只依賴于最初 $k-1$ 次試驗的結果，B_k 只依賴于第 k 次試驗，所以 A_k 及 B_k 是相互独立的，从而 $\mathbf{P}\{A_kB_k\}=\mathbf{P}\{A_k\}\mathbf{P}\{B_k\}=\frac{1}{2}\mathbf{P}\{A_k\}$. 故 $\mathbf{P}\{B\}=\sum\mathbf{P}\{A_kB_k\}=\frac{1}{2}\sum\mathbf{P}\{A_k\}=\frac{1}{2}$. 这就說明在这个长策下第一次賭博中出現“正面”的概率为$\frac{1}{2}$，同理,对于以后的各次賭博,这个結果也成立.

还要証明的是：要賭的那一串試驗是独立的. 这就意味着在第一第二两次賭博中錢币都出現“正面”的概率为$\frac{1}{4}$（以此类推到其它的組合以及此后的試驗）. 为了驗証这个論断,令 A_k^*表示“第二次打賭在第 k 次試驗”的事件. 令 E 表示“头两次賭博都出現‘正面’”的事件, E 是一切事件 $A_jB_jA_k^*B_k$ 的并，这里 $j<k$（如 $j\geqslant k$，則 A_j, A_k^* 互斥,从而 $A_jA_k^*=0$）. 因此

$$\mathbf{P}\{E\} = \sum_{j=1}^{\infty} \sum_{k=j+1}^{\infty} \mathbf{P}\{A_j B_j A_k^* B_k\}. \tag{2.2}$$

同样，我們可以看出，对固定的 j，$k > j$，事件 B_k（"第 k 次試驗出現正面"）与事件 $A_j B_j A_k^*$（它仅依賴于最初 $k-1$ 次試驗的結果）是相互独立的．因此

$$\mathbf{P}\{E\} = \frac{1}{2} \sum_{j=1}^{\infty} \sum_{k=j+1}^{\infty} \mathbf{P}\{A_j B_j A_k^*\}$$

$$= \frac{1}{2} \sum_{j=1}^{\infty} \mathbf{P}\{A_j B_j\} \sum_{k=j+1}^{\infty} \mathbf{P}\{A_k^* \mid A_j B_j\} \tag{2.3}$$

[参看第五章(1.8)]．但是，不管第一次賭是在哪次干的，也不管它的結果如何，賭博总是繼續下去的．这就是說：第二次賭博迟早会发生的，即对于已知具有 $\mathbf{P}\{A_j B_j\} > 0$ 的 $A_j B_j$，将"第二次賭博在第 k 次試驗发生"的条件概率对 k 加起来应为 1．所以(2.3)中的第二个級数之和为 1，同时我們早已知道 $\sum \mathbf{P}\{A_j B_j\} = \frac{1}{2}$，故 $\mathbf{P}\{E\} = \frac{1}{4}$．这就是我們所要証明的．对其它的組合可用类似的論証．

注意：如果賭者可以任意改变賭注的話，情况就不一样了．对于依賴累积赢得的长策，就存在所謂有利策略，而且賭博与策略有关．关于这一点，我們将在第十四章第 2 节里再来討論．

3. 波雷尔-康特立引理

有两个关于試驗的无穷序列的引理是經常要用到的，所以应該特别加以注意．我們仅仅对伯努利試驗来陈述这两个引理，但是，对于普遍的情况，它們还是可以应用的．

我們还是考虑一个无穷的伯努利試驗序列．令 $A_1, A_2, \cdots$ 表示事件的无穷序列，其中每一事件只依賴于有限次試驗．换言之，我們假定对于每一个 A_k，对应地存在一个 n_k，使得 A_k 是在最初的 n_k 次伯努利試驗的样本空間中的事件．令

$$a_k = \mathbf{P}\{A_k\}. \tag{3.1}$$

(例如，A_k 代表"在 $2k$ 次試驗完了时出現一个最少接連 k 次成功的連貫"的事件，則 $n_k = 2k$，而 $a_k = p^k$.)

对于字母 S 和 F 的某个特定的无穷序列，都可以确定它属于 $\{A_k\}$ 中的 0 个或 1 个或 2 个……或者无穷多个．这意味着我們可以談到 U_r，它是使事件族 $\{A_k\}$ 中有多于 r 个发生的无穷試驗序列．而 U_∞ 則表示使事件族 $\{A_k\}$ 中有无穷多个发生的无穷試驗序列．事件 U_r 只能在无穷的样本空間中定义，而它的概率是 $\mathbf{P}\{U_{n,r}\}$ 的极限，此处 $\mathbf{P}\{U_{n,r}\}$ 为 n 次試驗中有多于 r 个 A_k 发生的概率．最后，$\mathbf{P}\{U_\infty\} = \lim \mathbf{P}\{U_r\}$，此极限是存在的，因为当 r 增加时，$\mathbf{P}\{U_r\}$ 下降．

引理 1. 如果 $\sum a_k$ 收斂，則只有有限个 A_k 发生的概率为 1．更精确地說：就是要求，当 r 充分大时，$\mathbf{P}\{U_r\} < \varepsilon$．或者說：对每个 $\varepsilon > 0$ 可以找到这样一个整数 r，使得对每个 n，n 次試驗中 $A_{r+1}, A_{r+2}, \cdots$ 里的一个或多个发生的概率小于 ε．

証．由于級数 $\sum a_k$ 收斂，故我們可以取定一个 r，使 $a_{r+1} + a_{r+2} + \cdots < \varepsilon$．不失其普遍性，我們可以假設 A_k 是如此排列的，使得 $n_1 \leqslant n_2 \leqslant n_3 \leqslant \cdots$．令 N 为使 $n_N \leqslant n$ 成立的最后一个下标．于是 $A_1, \cdots, A_N$ 都在 n 次試驗的样本空間中定义了．而引理的結論是要求：在 $A_{r+1}, A_{r+2}, \cdots, A_N$ 中有一个或多个发生的概率小于 ε．这个結論是成立的，因为从第一章的不等式(6.6)我們有

$$\mathbf{P}\{A_{r+1} \cup A_{r+2} \cup \cdots \cup A_N\} \leqslant a_{r+1} + a_{r+2} + \cdots + a_N < \varepsilon. \quad (3.2)$$

至此，引理証毕．

上述引理，只有在 A_k 相互独立的条件下，才有一个圓滿的逆命題．当試驗串分为一些互不重迭的段落，而 A_k 只依賴于第 k 个段落中的試驗时，独立性条件就成立了．(例如：A_k 表示"第 k 个千次試驗中产生的成功次数多于 600"的事件．)

引理 2. 如果 A_k 相互独立，而且 $\sum a_k$ 发散，則有无穷个 A_k 发生的概率为 1．換言之，对任意一个 r，在 n 次試驗中多于 r 个 A_k 发生的概率随 $n \to \infty$ 而趋于 1．

証. 象在引理 1 中的証明一样，我們令 $A_1, A_2, \cdots, A_N$ 都定义在 n 次試驗的样本空間中．因为假定了它們相互独立，所以它們全不发生的概率是 $(1-a_1)(1-a_2)\cdots(1-a_N)$．但当 $0 < x < 1$ 时，$1-x < e^{-x}$，因此 $(1-a_1)(1-a_2)\cdots(1-a_N) < e^{-(a_1+a_2+\cdots+a_N)}$．因为 $\sum a_k$ 发散，故随着 N 的增大，$e^{-(a_1+a_2+\cdots+a_N)}$ 趋于零．此即証明了在 $\{A_k\}$ 中至少有一个发生的概率为 1．

其次，我們把 $\{A_k\}$ 分成两个子序列 $\{A_k^*\}$ 和 $\{A_k^{**}\}$，使两个級数 $\sum\mathbf{P}\{A_k^*\}$ 和 $\sum\mathbf{P}\{A_k^{**}\}$ 都发散．把上面的結論应用到这两个子序列上去，我們发現至少有一个 A_k^* 及一个 A_k^{**} 发生的概率为 1．因此 $\{A_k\}$ 中至少有两个发生的概率为 1．再把这个事实应用到 $\{A_k^*\}$ 和 $\{A_k^{**}\}$ 上去，則可以看出，$\{A_k\}$ 至少有 4 个发生的概率为 1．依此类推，郎可証明我們的引理．

例. 在一个伯努利試驗序列中，花样 SFS 出現无穷多次的概率为何？令 A_k 表示在第 $k, k+1, k+2$ 次出現花样 SFS．A_k 显然不是独立的，但是序列 $A_1, A_4, A_7, A_{10}, \cdots$ 只包含相互独立的事件．（因为其中任意两个事件都不依賴于同一次試驗的結果．）由于 $a_k = p^2q$ 是不依賴于 k 的，故級数 $a_1 + a_4 + a_7 + \cdots$ 发散，从而花样 SFS 无穷次地发生的概率为 1．显然，类似的論証可以应用到任何一个花样上去．

4. 强大数定律

概率的直观概念基于希望下述事实的成立：若 $\mathbf{S}_n$ 是伯努利試驗序列中前 n 次試驗的成功次数，則

$$\frac{\mathbf{S}_n}{n} \to p. \tag{4.1}$$

在抽象理論中，这不能对于每一个試驗序列都成立．事实上，在我們的样本空間中包含着这样一个点：它在无穷的試驗序列中每次都出現成功．对它来說：$\mathbf{S}_n/n = 1$，因此，(4.1)就不对了．然而，我們可以証明(4.1)概率为 1 地成立，于是(4.1)不成立的情形是可以忽略的例外．

注意： 在这里我們所討論的命題要比弱大数定律［第六章(4.2)］強得多．后者只是說，对每个充分大的 n 来說，平均数 $\mathbf{S}_n/n$ 与 p 差不多，但它并沒說当 n 增加时，$\mathbf{S}_n/n$ 一直停留在 p 的固定的邻域內．在 n 次附加的試験中，事件 $\mathbf{S}_{n+1}/(n+1)<p-\varepsilon$，或 $\mathbf{S}_{n+2}/(n+2)<p-\varepsilon$，$\cdots$，或 $\mathbf{S}_{2n}/2n<p-\varepsilon$ 中至少有一件发生仍是可能的，因为它是很多个小概率之和，而我們只知道每个个別的概率很小而已．但現在我們可以証明： $\mathbf{S}_n/n-p$ 概率为 1 变得很小，并且**保持**很小．

强大数定律． 对于每个 $\varepsilon>0$，在事件

$$\left|\frac{\mathbf{S}_n}{n}-p\right|>\varepsilon \tag{4.2}$$

中，只有有限个发生的概率为 1． 这个結論蘊含了“(4.1)概率为 1 地成立”． 在有限的样本空間中，它断言： 对每一組 $\varepsilon>0$，$\delta>0$，有一个 r，使得对一切 v 下面 v 个不等式：

$$\left|\frac{\mathbf{S}_{r+k}}{r+k}-p\right|<\varepsilon,\qquad k=1,2,\cdots,v, \tag{4.3}$$

同时成立的概率大于 $1-\delta$．

証． 我們将証明一个更強的命題．令 A_k 表示事件

$$|\mathbf{S}_k^*|=\left|\frac{\mathbf{S}_k-kp}{(kpq)^{\frac{1}{2}}}\right|>(2a\log k)^{\frac{1}{2}},\quad a>1. \tag{4.4}$$

于是由第七章公式(5.2)，至少对所有充分大的 k

$$\mathbf{P}\{A_k\}<e^{-a\log k}=\frac{1}{k^a}, \tag{4.5}$$

从而 $\sum\mathbf{P}\{A_k\}$ 收斂，因此由前一节的引理 1 可以保証： 只有有限个(4.4)型的不等式成立的概率为 1．另一方面，若(4.2)成立，则

$$\left|\frac{\mathbf{S}_n-np}{(npq)^{\frac{1}{2}}}\right|>\frac{\varepsilon}{(pq)^{\frac{1}{2}}}\cdot n^{\frac{1}{2}}, \tag{4.6}$$

而当 n 充分大时，右端大于 $(2a\log n)^{\frac{1}{2}}$． 因此无穷多个(4.2)型的不等式成立就蘊含了无穷多个 A_k 发生． 所以它（无穷多个(4.2)成立）的概率为零．这就証明了我們的強大数定律．

强大数法则最初为康特立所陈述，其先波雷尔及豪司道夫(Hausdorff)討論了一些特殊情况。正如弱大数定律一样它只是随机变量一般定理中的极为特别的情形。把我們的定理与长策之不可能性联系起来、不仅对原来的試驗序列而且对于按照第 2 节中的规则所得到的一切子序列来説，大数定律都蘊含了极限(4.1)的存在性。因而这两个定理一起描述了随机性的基本性質，这些性質都是概率的直观概念所固有的。它們的重要性曾由馮·米賽斯特别加以强調过。

5. 迭对数法則

象第七章一样，让我們再引进 n 次試驗中正則化的成功次数：

$$\mathbf{S}_n^* = \frac{\mathbf{S}_n - np}{(npq)^{\frac{1}{2}}}. \tag{5.1}$$

拉普拉斯极限定理断言： $\mathbf{P}\{\mathbf{S}_n^* > x\} \sim 1 - \Phi(x)$. 因此对于 n 的每一个特定的值来説，$\mathbf{S}_n^*$ 取大值的概率是不大的。但是，直观上看来很清楚：在漫长的試驗序列中，$\mathbf{S}_n^*$ 或迟或早地会取任意大的值。$\mathbf{S}_n^*$ 的中間值是最概然的，但是它的最大值也在慢步地增长着。增长的速率如何？在强大数定律的証明过程中，我們曾經断言，从(4.5)可推出对每一个 $a > 1$ 和全部充分大的 n、不等式 $\mathbf{S}_n^* < (2a \log n)^{\frac{1}{2}}$ 概率为 1 地成立。这就給出了 $\mathbf{S}_n^*$ 起伏的上界，但是这个上界不好。为了看出这点，让我們把这个結論应用到子序列 $\mathbf{S}_2^*, \mathbf{S}_4^*, \mathbf{S}_8^*, \mathbf{S}_{16}^*, \cdots$ 上；即我們定义事件 A_k 为 $\mathbf{S}_{2^k}^* \geq (2a \log k)^{\frac{1}{2}}$. 現在因为不等式(4.5)蘊含 $\mathbf{S}_{2^k}^* < (2a \log k)^{\frac{1}{2}}$ 对 $a > 1$ 和一切充分大的 k. 但是，对 $n = 2^k$，我們有 $\log k \sim \log \log n$，因此我們断言：对于每一个 $a > 1$ 和全部形如 $n = 2^k$ 的 n，不等式

$$\mathbf{S}_n^* < (2a \log \log n)^{\frac{1}{2}} \tag{5.2}$$

对某个 k 以后全成立。所以合适地猜测，应該是(5.2)对一切充分大的 n 都成立，事实上，它是迭对数法則的一部分。这个定

理[1]断言：$(2\log\log n)^{\frac{1}{2}}$是在下述意义下的精确的上界：对于每一个 $a>1$，不等式(5.2)的反不等式对无穷多个 n 成立.

定理. 概率为 1 地有

$$\limsup_{n\to\infty}\frac{\mathbf{S}_n^*}{(2\log\log n)^{\frac{1}{2}}}=1. \tag{5.3}$$

这意思是説：如果 $\lambda>1$，則事件

$$\mathbf{S}_n>np+\lambda(2npq\log\log n)^{\frac{1}{2}} \tag{5.4}$$

之中只有有限个发生的概率是 1；如果 $\lambda<1$，則(5.4)对于无穷多个 n 成立的概率是 1.

由于对称性，方程(5.3)蘊含了

$$\liminf_{n\to\infty}\frac{\mathbf{S}_n^*}{(2\log\log n)^{\frac{1}{2}}}=-1. \tag{5.3a}$$

証. 我們先从以下两个預备性的附注入手.

(1) 存在这样一个常数 $c>0$，它依賴于 p 而不依賴于 n，使得对一切的 n 都有

$$\mathbf{P}\{\mathbf{S}_n>np\}>c. \tag{5.5}$$

事实上，考察二項分布，得知(5.5)的左边永不为零，而且由拉普拉斯极限定理得知当 $n\to\infty$ 时它趋于$\frac{1}{2}$. 所以它总有大于 0 之下界. 这就証明了我們的附注.

(2) 我們需要下面的引理. 令 x 固定，且令 A 表示以下事件：至少有一个 k, $k\leqslant n$，使

$$\mathbf{S}_k-kp>x. \tag{5.6}$$

于是

$$\mathbf{P}\{A\}\leqslant\frac{1}{c}\mathbf{P}\{\mathbf{S}_n-np>x\}. \tag{5.7}$$

为了証明引理，令 A_ν 表示“(5.6)对 $k=\nu$ 时成立而对 $k=1, 2, \cdots, \nu-1$ 时不成立”(此处 $1\leqslant\nu\leqslant n$)的事件. $A_1, A_2, \cdots,$

---

1) 参看[60]. 这个发現中的部分結果曾先由其他的作者得到过. 現在的証明是为了可以直接推广到一般的随机变量上的情形.

A_n 是互斥的．同时它們的并是 A．因此有

$$\mathbf{P}\{A\} = \mathbf{P}\{A_1\} + \mathbf{P}\{A_2\} + \cdots + \mathbf{P}\{A_n\}. \tag{5.8}$$

其次对 $\nu < n$，令 U_ν 表示“在第 $\nu+1, \nu+2, \cdots, n$ 次試驗中成功的总次数超过 $(n-\nu)p$”的事件． 若两个事件 A_ν, U_ν 都发生，则 $\mathbf{S}_n > \mathbf{S}_\nu + (n-\nu)p > np + x$，又因为 $\{A_\nu U_\nu\}$ 是互斥的，所以

$$\begin{aligned}&\mathbf{P}\{\mathbf{S}_n - np > x\}\\ \geqslant\ &\mathbf{P}\{A_1U_1\} + \mathbf{P}\{A_2U_2\} + \cdots + \mathbf{P}\{A_{n-1}U_{n-1}\} + \mathbf{P}\{A_nU_n\}.\end{aligned} \tag{5.9}$$

因为 A_ν 只依赖于最初那 ν 次試驗，U_ν 只依赖于其后的 $n-\nu$ 次試驗．因此 A_ν, U_ν 是相互独立的，从而 $\mathbf{P}\{A_\nu U_\nu\} = \mathbf{P}\{A_\nu\}\mathbf{P}\{U_\nu\}$．从(5.5)中我們知道，$\mathbf{P}\{U_\nu\} > c$，且因 $c < 1$，故从(5.8)和(5.9)可得

$$\mathbf{P}\{\mathbf{S}_n - np > x\} \geqslant c\sum\mathbf{P}\{A_\nu\} = c\mathbf{P}\{A\}. \tag{5.10}$$

这就証明了(5.7)．

(3) 現在我們来証明定理中关于(5.4)式中 $\lambda > 1$ 的那个部分．令 γ 表示这样一个数，使得

$$1 < \gamma < \lambda^{\frac{1}{2}}, \tag{5.11}$$

而且令 n_r 表示最靠近 $\gamma^r (r = 1, 2, \cdots)$ 的整数．令 B_r 表示以下事件：至少对一个 n，$n_r \leqslant n < n_{r+1}$

$$\mathbf{S}_n - np > \lambda(2n_r pq \log\log n_r)^{\frac{1}{2}} \tag{5.12}$$

成立． 很显然只有当无穷多个 B_r 发生时(5.4)才能对无穷多个 n 成立．应用第一个波雷尔-康特立引理，我們看出只需証明

$$\sum\mathbf{P}\{B_r\}\text{收斂} \tag{5.13}$$

就够了．由不等式(5.7)可知

$$\begin{aligned}\mathbf{P}\{B_r\} &\leqslant c^{-1}\mathbf{P}\{\mathbf{S}_{n_{r+1}} - n_{r+1}p > \lambda(2n_r pq \log\log n_r)^{\frac{1}{2}}\}\\ &= c^{-1}\mathbf{P}\left\{\mathbf{S}^*_{n_{r+1}} > \lambda\left(2\frac{n_r}{n_{r+1}}\log\log n_r\right)^{\frac{1}{2}}\right\}.\end{aligned} \tag{5.14}$$

因为 $n_r/n_{r+1} \sim \gamma^{-1} > \lambda^{-\frac{1}{2}}$，所以对充分大的 r，有

$$\mathbf{P}\{B_r\} \leqslant c^{-1}\mathbf{P}\{\mathbf{S}^*_{n_{r+1}} > (2\lambda\log\log n_r)^{\frac{1}{2}}\}. \tag{5.15}$$

再利用德模佛-拉普拉斯极限定理(第七章公式(5.2))，可知，对相

当大的 r，有

$$\mathbf{P}\{B_r\} \leqslant c^{-1}e^{-\lambda\log\log n_r} = \frac{1}{c(\log n_r)^\lambda} \sim \frac{1}{c(r\log\gamma)^\lambda}. \quad (5.16)$$

因为 $\lambda > 1$，故(5.13)的結論得証.

(4) 最后，我們来証明关于(5.4)中的 $\lambda<1$ 的那一部分. 这时我們选取 γ 为很大的整数，使得

$$\frac{\gamma-1}{\gamma} > \eta > \lambda, \quad (5.17)$$

其中 η 是在以后要决定的一个常数. 令 $n_r = \gamma^r$，由于第二个波雷尔-康特立引理只能对独立事件才能应用，故我們引进

$$\mathbf{D}_r = \mathbf{S}_{n_r} - \mathbf{S}_{n_{r-1}}; \quad (5.18)$$

$\mathbf{D}_r$ 是从第 n_{r-1} 次以后到第 n_r 次（包括第 n_r 次）試驗中成功的总数；对 $\mathbf{D}_r$ 我們有二項分布 $b(k; n, p)$，其中 $n = n_r - n_{r-1}$. 令 A_r 为事件

$$\mathbf{D}_r - (n_r - n_{r-1})p > \eta(2pqn_r\log\log n_r)^{\frac{1}{2}}. \quad (5.19)$$

往証无穷多个 A_r 概率为 1 地发生. 由于各个不同的 A_r 依賴于互不重迭的試驗段落，(即 $n_{r-1} < n \leqslant n_r$，亦即 A_r 只依賴第 $n_{r-1}+1$ 到第 n_r 次試驗.)故 $\{A_r\}$ 是相互独立的. 因此，根据第二个波雷尔-康特立引理，只須証明：$\sum\mathbf{P}\{A_r\}$ 发散就够了. 但

$$\mathbf{P}\{A_r\}$$
$$= \mathbf{P}\left\{\frac{\mathbf{D}_r - (n_r - n_{r-1})p}{\{(n_r - n_{r-1})pq\}^{\frac{1}{2}}} > \eta\left(2\frac{n_r}{n_r - n_{r-1}}\log\log n_r\right)^{\frac{1}{2}}\right\}. \quad (5.20)$$

由(5.17)可知 $n_r/(n_r - n_{r-1}) = \gamma/(\gamma-1) < \eta^{-1}$. 因此

$$\mathbf{P}\{A_r\} \geqslant \mathbf{P}\left\{\frac{\mathbf{D}_r - (n_r - n_{r-1})p}{\{(n_r - n_{r-1})pq\}^{\frac{1}{2}}} > (2\eta\log\log n_r)^{\frac{1}{2}}\right\}. \quad (5.21)$$

再利用第七章(5.2)的估計，我們发現对相当大的 r

$$\mathbf{P}\{A_r\} > \frac{1}{2\eta\log\log n_r}e^{-\eta\log\log n_r} = \frac{1}{2\eta(\log\log n_r)(\log n_r)^\eta}. \quad (5.22)$$

因为 $n_r = \gamma^r$，而且 $\eta < 1$，故对充分大的 r 我們有 $\mathbf{P}\{A_r\} > \frac{1}{r}$，

这就証明了級数 $\sum \mathbf{P}\{A_r\}$发散.

証明的最后一步是要驗証(5.18)中的 $\mathbf{S}_{n_{r-1}}$ 可以略去. 从我們已經証明了的定理的第一部分可知,对每一个 $\varepsilon > 0$ 都可以找到这样的一个 N, 使得对一切 $r > N$

$$|\mathbf{S}_{n_{r-1}} - n_{r-1}p| < 2(2pqn_{r-1}\log\log n_{r-1})^{\frac{1}{2}} \tag{5.23}$$

的概率大于等于 $1-\varepsilon$. 現在选取 η 非常接近于1,使得

$$1-\eta < \left(\frac{\eta-\lambda}{2}\right)^2. \tag{5.24}$$

于是从(5.17)可得

$$4n_{r-1} = 4\frac{n_r}{\gamma} < n_r(\eta-\lambda)^2, \tag{5.25}$$

因此(5.23)就蘊含

$$\mathbf{S}_{n_{r-1}} - n_{r-1}p > -(\eta-\lambda)(2pqn_r\log\log n_r)^{\frac{1}{2}}. \tag{5.26}$$

把(5.26)加到(5.19)上去,我們得到(5.4)对 $n = n_r$ 成立. 这就推出这个不等式对无穷多个 r 成立的概率大于等于 $1-\varepsilon$, 这就完成了我們的証明.

伯努利試驗的迭对数法則是一个普遍定理的特殊情形, 这一个普遍的定理是由柯尔莫格洛夫[61]所首先提出的. 現在有可能提出更強的定理(参看問題7及問題8).

6. 用数論的語言解释

令 x 为区間 $0 \leqslant x < 1$ 中的一个实数,且令

$$x = 0.a_1a_2a_3\cdots \tag{6.1}$$

是它的十进位小数的展开式(于是每个 a_j 代表数字 $0, 1, \cdots, 9$ 中的某一个). 除了 $a/10^n$ 型的数以外,上面的展开式是唯一的. 因为 $a/10^n$ 旣可使其展开式表为有限多位(包含无穷多个0)又可使其展开式包含无穷多个9. 为了避免混乱起見, 我們决定不用后一种形式.

十进位小数的展开式可以与一个具有 $p = \frac{1}{10}$ 的伯努利試驗联系起来, 数字0代表成功, 其它的9个数字代表失敗. 如果在

(6.1) 中以 S 代替 0，以 F 代替其它的数字，(6.1) 就代表了具有 $p=\frac{1}{10}$ 的伯努利試驗无穷序列的一个可能結果. 反过来,任何一个由 S 和 F 所組成的序列都可以由某个 x 的十进位小数的展开式按上面的方式得之. 按照这种方法,伯努利試驗的样本空間中每个事件都可以用 x 的某个集合来表示. 例如:“第 n 次試驗为成功”的事件可以用第 n 位小数为 0 的那些 x 来表示. 这些 x 就构成了 10^{n-1} 个区間,每个区間之长为 10^{-n},它們的总长为 $\frac{1}{10}$, 这就是这一事件的概率. 每一个长为 n 的有限的样本序列对应于某些区間的集合,例如,序列 SFS 对应于下面 9 个区間: $0.01\leqslant x<0.011$, $0.02\leqslant x<0.021, \cdots$, $0.09\leqslant x<0.091$. 每一个这样的样本序列的概率等于其所对应于 x 軸上之区間集合的总长度. 更复杂一些的事件常常可以用有限个样本序列的并来表示，从而計算前者的概率可以按照与 x 軸上常用的勒貝格 (Lebesgue) 測度相同的加法規律来进行. 因此,我們的概率永远可以与 x 軸上对应的点集的測度一致. 于是我們可以把具有 $p=\frac{1}{10}$ 的伯努利試驗的一切极限定理翻譯成关于十进位小数展开式的定理. 术語“概率为 1 地”等价于“几乎一切 x”或者“几乎处处”.

我們曾經考虑过 n 次試驗中成功的次数 $\mathbf{S}_n$ 这一随机变量. 为了方便起見，在这里我們着重指出下面的事实: $\mathbf{S}_n$ 是样本点的函数,我們用 $\mathbf{S}_n(x)$ 代表 x 的前 n 位小数中 0 的个数. 显然, $\mathbf{S}_n(x)$ 是 x 的函数,其图形为阶梯多边形,它的不連續点只能在 $a/10^n$ 型的点上,其中 a 为整数. 比值 $\mathbf{S}_n(x)/n$ 称为 x 的前 n 位小数中零的頻率.

在通常測度論的語言里,弱大数定律所断言的是: $\mathbf{S}_n(x)/n\rightarrow\frac{1}{10}$ 按測度收斂. 而強大数定律所断言的是: $\mathbf{S}_n(x)/n\rightarrow\frac{1}{10}$ 几乎处处收斂. 辛欽(А. Я. Хинчин)的迭对数法則就是說

$$\limsup_{n\to\infty} \frac{\mathbf{S}_n(x) - \frac{n}{10}}{(n\log\log n)^{\frac{1}{2}}} = (0.3)2^{\frac{1}{2}} \tag{6.2}$$

对几乎一切 x 成立. 它回答了一系列論文[62,63]中所研究过的一个問題. 这个結果的更进一步的改善,請参看問題 7 及問題 8 .

代替数字 0 ,我們可以考慮用其它的数字. 那么強大数定律可以解释为: 对几乎一切 x, 这 $(0, 1, \cdots, 9)$ 十个数字中每一数字的频率都趋于 $\frac{1}{10}$. 如果把十进位系統的基"10"换为其它的基,就可得出类似的定理. 这个事实曾为波雷尔(1909)所发現, 且往往表述为: 几乎一切的数都是"正常的".

7. 問　　題

1.求一个整数 β, 使得在擲骰子时"連續出現三次么点的連貫"先于"长为 β 的非么点的連貫"的出現,而出現的概率大約为 $\frac{1}{2}$.

2.考虑具有三个可能結果 A, B, C 的相互独立的重复試驗, A, B, C 所对应的概率分别为 p, q, r $(p+q+r=1)$. 求"接連出現 α 个 A 的連貫"在"接連出現 β 个 B 的連貫"之前的概率.

3 (續上題). 求长为 α 的 A-連貫发生在一个长为 β 的 B-連貫或在一个长为 γ 的 C-連貫之前的概率.

4.在伯努利試驗序列中, 令 A_n 表示在第 2^n 次到第 2^{n+1} 次試驗中接連出現 n 次成功的事件. 如果 $p \geq \frac{1}{2}$,則有无穷多个 A_n 发生的概率为 1; 如果 $p < \frac{1}{2}$,則只有有限个 A_n 发生的概率为 1.

5[1].令 $\mathbf{N}_n$ 为在第 n 次試驗开始的成功連貫的长度 (即是,如果第 n 次試驗的結果是 F, 則 $\mathbf{N}_n = 0$, 等等). 証明

$$\limsup \frac{\mathbf{N}_n}{\operatorname{Log} n} = 1 \tag{7.1}$$

的概率为 1, 其中 Log 表示以 $1/p$ 为底的对数.

提示: 考虑一个长度超过 $a \operatorname{Log} n$ 的成功連貫在第 n 次試驗之后的事件

1) 从奈依曼(D. J. Newman)的一篇稿件中提出的.

A_n. 对 $a>1$，計算是非常順利的. 对 $a<1$，考虑試驗次数的一个子序列 $a_1, a_2, \cdots$，其中 a_n 是与 $n \operatorname{Log} n$ 非常接近的整数.

6. 从迭对数法則我們可以推出：有无穷多个 n 使 $\mathbf{S}_k (n<k<17n)$ 全是正的概率为 1.（注意：应用第三章的結果，可以証明更强的結論.）

7. 令 $\phi(t)$ 为正的单調上升函数，且令 n_r 为最接近 $e^{r/\log r}$ 的整数，如果

$$\sum \frac{1}{\phi(n_r)} e^{-\frac{1}{2}\phi^2(n_r)} \tag{7.2}$$

收斂，則概率为 1 地只有有限个 n 使

$$\mathbf{S}_n > np + (npq)^{\frac{1}{2}}\phi(n) \tag{7.3}$$

成立. [注意：不妨假設 $\phi(n) < 10(\log\log n)^{\frac{1}{2}}$，因为对于更大的 $\phi(n)$，迭对数法則足够对付这个問題.]

8. 証明[1)]級数(7.1)收斂当且仅当

$$\sum \frac{\phi(n)}{n} e^{-\frac{1}{2}\phi^2(n)} \tag{7.4}$$

收斂. [提示：归并对应于 $n_{r-1} < n < n_r$ 的那些項，并注意 $n_r - n_{r-1} \sim n_r(1 - 1/\log r)$；此外(7.3)只有当 $\phi^2(n) > 2\log\log n$ 时才能收斂.]

---

1) 問題 7 問題 8 合在一起証明了：在不等式(7.4)收斂的情形下，概率为 1 地只有有限个 n 使不等式(7.3)成立. 反过来，如果(7.4)发散，則概率为 1 地有无穷多个 n 使不等式(7.3)成立. 其逆是很难証明的，参看[64]，其中对于任意随机变量的更一般的定理都証明了. 对于具有 $p=\frac{1}{2}$ 的伯努利試驗的特殊情形，請参閱[65]，迭对数法則是从 $\phi(t) = \lambda(2\log\log t)$ 的特别情形得出的.

第九章　随机变量;期望值

1. 随机变量

按照微积分教科书上的定义，如果对每一个实数 x 都有唯一的一个值 y 与之对应，则称 y 为实变量 x 的函数．这个定义可以推广到自变量 x 不是实数的情形．我們可以称两点間之距离是一对点的函数;三角形之周长为定义在三角形集合上的函数;序列 a_n 是定义在所有正整数上的函数;二项式系数 $\binom{x}{k}$ 为数对 (x, k) (其中第二个数 k 为非負整数)的函数．同样,我們可以称 n 次伯努利試驗的成功次数 $\mathbf{S}_n$ 为样本空間上的函数; 这个空間由 2^n 个样本点組成,且每一个点都对应于一个数 $\mathbf{S}_n$.

定义在样本空間上的函数就称为随机变量．在前面各章里面我們一直用着随机变量的概念,只是沒用这个术語罢了．举几个典型例子：打桥牌时手中爱司的个数是随机变量；n 个人中生日相同的人数是随机变量；n 次伯努利試驗中成功的連貫数是随机变量．在每一种情形都有唯一的一个规則使每一个样本点与一个数 $\mathbf{X}$ 联系起来．概率的古典理論主要是致力于賭徒的贏得的研究．而賭徒所获之贏得是一个随机变量．事实上,每一个随机变量都可以解释为一个眞实的或者假想的賭徒在适当賭博中所获之贏得．在物理系統中扩散时质点的位置、能量、温度等等都是随机变量,但是它們是定义在非离散的样本空間中，因而,对于它們的研究将放在以后．在离散的样本空間的情况下,我們眞正可以把任一随机变量 $\mathbf{X}$ 列表表示，即用某种次序把样本空間中的所有的点列出而且与每一个点相对应的 $\mathbf{X}$ 的值也都列出来．

随机变量这一术語有些模糊，而說随机函数则更合适一些

[其自变量为样本空間中的点，亦即一次实驗的結果].

令 $\mathbf{X}$ 为随机变量，$x_1, x_2, x_3, \cdots$ 为 $\mathbf{X}$ 所取之值[1]（在下面大多数情形中，x_j 将为整数）．所有的那些样本点，使得在其上 $\mathbf{X}$ 取固定值 x_j 者构成事件：$\mathbf{X}=x_j$；它的概率以 $\mathbf{P}\{\mathbf{X}=x_j\}$ 表示．下列函数：

$$\mathbf{P}\{\mathbf{X}=x_j\}=f(x_j) \quad (j=1, 2, \cdots) \tag{1.1}$$

叫作随机变量 $\mathbf{X}$ 的(概率)分布[2]．显然

$$f(x_j)\geqslant 0, \quad \sum f(x_j)=1. \tag{1.2}$$

如果应用这些术語，我們可以說伯努利試驗中的成功次数 $\mathbf{S}_n$ 是一个具有概率分布 $\{b(k;n, p)\}$ 的随机变量，而一直到第一次成功出現（包含这次試驗）所需之試驗次数是一个具有概率分布 $\{q^{k-1}p\}$ 的随机变量．

現在我們考虑定义在同一个样本空間中的两个随机变量 $\mathbf{X}$ 和 $\mathbf{Y}$，其所取之値分别記作 $x_1, x_2, \cdots$ 和 $y_1, y_2, \cdots$；且其对应的概率分布为 $\{f(x_j)\}$ 和 $\{g(y_k)\}$．同时满足 $\mathbf{X}=x_j$ 和 $\mathbf{Y}=y_k$ 这两个条件的点的总体构成一个事件，其概率記作 $\mathbf{P}\{\mathbf{X}=x_j; \mathbf{Y}=y_k\}$．下列函数：

$$\mathbf{P}\{\mathbf{X}=x_j, \mathbf{Y}=y_k\}=p(x_j, y_k) \quad (j, k=1, 2, \cdots) \tag{1.3}$$

叫作 $\mathbf{X}, \mathbf{Y}$ 之联合概率分布．最好是用如表 1，2 所例示的双重記录表来表示它．(注意：行数与列数不一定要求相等)显見

$$p(x_j, y_k)\geqslant 0, \qquad \sum_{j,k} p(x_j, y_k)=1, \tag{1.4}$$

且对每一个固定的 j 有

---

1) 在标准的数学术語里，点集 $x_1, x_2, \cdots$ 称为 $\mathbf{X}$ 的值域．不幸的是，統計的文献里，把 $\mathbf{X}$ 的最大値和最小値之差称为値域．

2) 对一个离散的随机变量 $\mathbf{X}$ 来說，其概率分布为定义在 $\mathbf{X}$ 所能够取的值 x_j 的总体上的函数 $f(x_j)$，这个术語必須要与"分布函数"这一术語区别开来．分布函数是一个非降的函数，且当 $x\to-\infty$ 时它趋于 0；当 $x\to\infty$ 时，它趋于 1．$\mathbf{X}$ 之分布函数 $F(x)$ 用下式来定义：

$$F(x)=\mathbf{P}\{\mathbf{X}\leqslant x\}=\sum_{x_j\leqslant x} f(x_j),$$

最后的和是跑遍一切不超过 x 的 x_j．故一个随机变量的分布函数可以从其概率分布計算而得，反过来說也对．在这一卷里我們不討論一般的分布函数．

$$p(x_j, y_1) + p(x_j, y_2) + p(x_j, y_3) + \cdots$$
$$= \mathbf{P}\{\mathbf{X} = x_j\} = f(x_j); \tag{1.5}$$

表 1. 例(a)中$(\mathbf{N}, \mathbf{X}_1)$的联合分布

		$\mathbf{X}_1$ 0	1	2	3	$\mathbf{N}$ 的分布
	1	$2q$	0	0	q	$3q = \frac{1}{9}$
$\mathbf{N}$	2	$6q$	$6q$	$6q$	0	$18q = \frac{2}{3}$
	3	0	$6q$	0	0	$6q = \frac{2}{9}$
$\mathbf{X}_1$ 的分布		$8q$	$12q$	$6q$	q	

$$\mathbf{E}(\mathbf{N}) = \frac{19}{9}, \quad \mathbf{E}(\mathbf{N}^2) = \frac{129}{27}, \quad \mathrm{Var}(\mathbf{N}) = \frac{26}{81},$$
$$\mathbf{E}(\mathbf{X}_1) = 1, \quad \mathbf{E}(\mathbf{X}_1^2) = \frac{45}{27}, \quad \mathrm{Var}(\mathbf{X}_1) = \frac{2}{3},$$
$$\mathbf{E}(\mathbf{N},\mathbf{X}_1) = \frac{19}{9}, \quad \mathrm{Cov}(\mathbf{N},\mathbf{X}_1) = 0.$$

$\mathbf{N}$ 是 3 个球随机地放入三个盒中被装进球的盒的个数，$\mathbf{X}_1$ 是第一个盒中的球的个数. 簡記 $q = \frac{1}{27}$.

表 2. 例(a)中$(\mathbf{X}_1, \mathbf{X}_2)$的联合分布

		$\mathbf{X}_1$ 0	1	2	3	$\mathbf{X}_1$ 的分布
	0	q	$3q$	$3q$	q	$8q$
	1	$3q$	$6q$	$3q$	0	$12q$
$\mathbf{X}_2$	2	$3q$	$3q$	0	0	$6q$
	3	q	0	0	0	q
$\mathbf{X}_2$ 的分布		$8q$	$12q$	$6q$	q	

$$\mathbf{E}(\mathbf{X}_i) = 1, \quad \mathbf{E}(\mathbf{X}_i^2) = \frac{45}{27}, \quad \mathrm{Var}(\mathbf{X}_i) = \frac{2}{3},$$
$$\mathbf{E}(\mathbf{X}_1,\mathbf{X}_2) = \frac{2}{3}, \quad \mathrm{Cov}(\mathbf{X}_1,\mathbf{X}_2) = -\frac{1}{3}.$$

$\mathbf{X}_i$ 是 3 个球随机地放入三个盒中第 i 个盒中的球的个数. 簡記 $q = \frac{1}{27}$.

对每一个固定的 k 有

$$p(x_1, y_k) + p(x_2, y_k) + p(x_3, y_k) + \cdots \\ = \mathbf{P}\{\mathbf{Y} = y_k\} = g(y_k). \qquad (1.6)$$

换言之，把每行或每列的概率加起来，我們分别得到 $\mathbf{X}$ 或 $\mathbf{Y}$ 的概率分布．这两个分布叫作边緣分布；它們可以在表上显示出来，象表 1, 2 中那样．“边緣”这一形容詞是从双重記录表的外貌想出来的．而且当上下文中两个随机变量的联合分布及它們的单个（边緣）分布同时出現时，为了文体上的清晰，也应用“边緣”这一形容詞以示区别．严格地說：“边緣”这一形容詞总是多余的．

联合分布这一概念完全可以移植到多于两个随机变量的系統上去．

例． (a) 把 3 个球随机地放入 3 个盒内． 我們考虑一个形式上由表 1 及第一章例 (2.a) 所定义的 27 个点构成的样本空間，对每一个点我們賦以概率 1/27．令 $\mathbf{N}$ 为已放有球的盒的数目，令 $\mathbf{X}_i$ 为第 i 个盒中的球的个数，$i = 1, 2, 3$．这些都是形象性的描述．形式上看 $\mathbf{N}$ 这个函数在第 1–3 个样本点上取值 1；在第 4–21 个样本点上取值为 2；在第 22–27 个样本点上取値为 3． 因此，$\mathbf{N}$ 的概率分布为 $\mathbf{P}\{\mathbf{N} = 1\} = \frac{1}{9}$，$\mathbf{P}\{\mathbf{N} = 2\} = \frac{2}{3}$，$\mathbf{P}\{\mathbf{N} = 3\} = \frac{2}{9}$．$(\mathbf{N}, \mathbf{X}_1)$和$(\mathbf{X}_1, \mathbf{X}_2)$的联合分布由表 1 和表 2 給出．

(b) 骰子． 把一个均匀的骰子擲 n 次，令 $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ 分别为么点，二点，三点出現的次数．擲 n 次的結果么点出現 k_1 次；二点出現 k_2 次；三点出現 k_3 次，而其它各点出現 $n - k_1 - k_2 - k_3$ 次的概率 $p(k_1, k_2, k_3)$由第六章多項分布(9.2)所給出，其中 $p_1 = p_2 = p_3 = \frac{1}{6}$，$p_4 = \frac{1}{2}$，也即

$$p(k_1, k_2, k_3) = \frac{n!}{k_1!k_2!k_3!(n-k_1-k_2-k_3)!} \cdot 3^{n-k_1-k_2-k_3}6^{-n}. \qquad (1.7)$$

这就是 $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ 的联合分布．設 k_1, k_2 固定，把 (1.7) 对一切可能的 $k_3 = 0, 1, \cdots, n - k_1 - k_2$ 求和，再应用二項定理我們得到

$$p(k_1, k_2) = \frac{n!}{k_1!k_2!(n-k_1-k_2)!} \cdot 4^{n-k_1-k_2}6^{-n}. \qquad (1.8)$$

这就是 $(\mathbf{X}_1, \mathbf{X}_2)$ 的联合分布，但是，它又是 $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ 的三維分布的边緣分布．无需説，(1.8)可以由多項分布直接得出．把(1.8)再一次对全部的 $k_2 = 0, 1, \cdots, n-k_1$ 求和，我們得到 $\mathbf{X}_1$ 的分布，即是具有 $p = \frac{1}{6}$ 的二項分布．

(c) 抽样． 把一个具有 n 个元素的总体分为三类，其大小分别为 $n_1 = np_1$，$n_2 = np_2$，$n_3 = np_3$（其中 $p_1 + p_2 + p_3 = 1$）．設从总体中抽取一个大小为 r 的随机样本，并令 $\mathbf{X}_1$、$\mathbf{X}_2$ 分别为这个样本中第一类和第二类中的元素的个数． 如果抽样是有放回的，則 $\mathbf{P}\{\mathbf{X}_1 = k_1, \mathbf{X}_2 = k_2\}$ 由多項分布

$$f(k_1, k_2) = \frac{r!}{k_1!k_2!(r-k_1-k_2)!} p_1^{k_1}p_2^{k_2}p_3^{r-k_1-k_2} \qquad (1.9)$$

所給出[参看第六章公式(9.2)]．随机变量 $\mathbf{X}_i$ 具有二項分布 $b(k; n, p_i)$，如果抽样是无放回的，則 $\mathbf{P}\{\mathbf{X}_1 = k_1, \mathbf{X}_2 = k_2\}$ 由第二章(6.5)所給出，而 $\mathbf{X}_1$ 則具有第二章(6.1)的超几何分布．

(d) 随机抽样． 我們再来考虑前一个例子，只不过現在样本的大小 r 不是事先給定的，而是依賴于一个随机試驗的結果． 更确切地説，假定样本的大小服从普阿松分布：样本的大小为 r 的概率为 $p(r; \lambda) = e^{-\lambda}\lambda^r/r!$，而在样本的大小 r 給定后，$\mathbf{X}_1 = k_1$ 和 $\mathbf{X}_2 = k_2$ 的（条件）概率由 (1.9) 的 $f(k_1, k_2)$ 所給出． 因此，$(\mathbf{X}_1, \mathbf{X}_2)$的联合分布为

$$\mathbf{P}\{\mathbf{X}_1 = k_1, \mathbf{X}_2 = k_2\} = e^{-\lambda} \sum_{r=k_1+k_2}^{\infty} \lambda^r f(k_1, k_2)/r! =$$

$$= e^{-\lambda} \cdot \frac{(\lambda p_1)^{k_1}(\lambda p_2)^{k_2}}{k_1!k_2!} \sum_{k_3=0}^{\infty} \frac{(\lambda p_3)^{k_3}}{k_3!} =$$

$$= e^{-\lambda(1-p_3)} \frac{(\lambda p_1)^{k_1}(\lambda p_2)^{k_2}}{k_1!k_2!} \qquad (1.10)$$

或

$$\mathbf{P}\{\mathbf{X}_1 = k_1, \mathbf{X}_2 = k_2\} = p(k_1; \lambda p_1) p(k_2; \lambda p_2). \qquad (1.11)$$

对所有的 k_2 求和，我們看到 $\mathbf{X}_1$ 具有普阿松分布 $p(k;\ \lambda p_1)$。(第六章問題 27 說明同一个事实.) $(\mathbf{X}_1, \mathbf{X}_2)$ 的联合分布取边緣分布 $\{p(k;\ \lambda p_1)\}$ 和 $\{p(k;\ \lambda p_2)\}$ 的乘法表的形式. 这时我們就說 $\mathbf{X}_1$ 和 $\mathbf{X}_2$ 独立.

应用(1.3)的符号，则在 $\mathbf{X} = x_j$ 給定的条件下 $(f(x_j) > 0)$，事件 $\mathbf{Y} = y_k$ 发生的条件概率变为

$$\mathbf{P}\{\mathbf{Y} = y_k | \mathbf{X} = x_j\} = \frac{p(x_j, y_k)}{f(x_j)}. \tag{1.12}$$

为了方便起見，把(1.12)簡写为 $\mathbf{P}\{\mathbf{Y} = y_k | \mathbf{X}\}$；这就定义了給定 $\mathbf{X}$ 后，$\mathbf{Y}$ 的(条件)分布. 从表 1 和表 2 中一看便知：一般地說，条件概率(1.12)与 $g(y_k)$ 是不同的. 这表明从 $\mathbf{X}$ 的值我們可以对 $\mathbf{Y}$ 的值作一些推測，反过来說也对；这两个随机变量是(統計地)不独立的. 当 $\mathbf{Y}$ 是 $\mathbf{X}$ 的函数，也就是当 $\mathbf{X}$ 的値唯一决定 $\mathbf{Y}$ 时，$\mathbf{X}$ 与 $\mathbf{Y}$ 之間存在着最強度的依賴性. 例如把一个錢币扔 n 次，令 $\mathbf{X}$ 和 $\mathbf{Y}$ 分别为"正面"及"反面"所出現的次数，则 $\mathbf{Y} = n - \mathbf{X}$. 同样，当 $\mathbf{Y} = \mathbf{X}^2$ 时，我們可以由 $\mathbf{X}$ 来計算 $\mathbf{Y}$. 这就意味着，在联合分布中每一行除了一項以外其他的都是 0. 另一方面，如果对一切組合 x_j, y_k 都有 $p(x_j, y_k) = f(x_j)g(y_k)$，则事件 $\mathbf{X} = x_j$ 与 $\mathbf{Y} = y_k$ 是相互独立的；且联合分布取乘法表的形式. 这时我們称随机变量 $\mathbf{X}, \mathbf{Y}$ 是相互独立的. 特别地联系于独立的試驗就会出現相互独立的随机变量. 例如：把一顆骰子擲两次其所得之点数是相互独立的随机变量. 不同性质的一个例子可見例(d).

注意：$\mathbf{X}$ 和 $\mathbf{Y}$ 的联合分布决定 $\mathbf{X}, \mathbf{Y}$ 的分布. 但是从 $\mathbf{X}, \mathbf{Y}$ 的边緣分布不能算出 $\mathbf{X}$ 和 $\mathbf{Y}$ 的联合分布. 如果两个随机变量 $\mathbf{X}, \mathbf{Y}$ 有相同的分布，则它們可能相互独立也可能不相互独立. 例如在表 2 中的两个随机变量 $\mathbf{X}_1, \mathbf{X}_2$ 不是相互独立的然而却具有相同的分布.

所有这些概念都可以应用到多于两个随机变量的情形. 我們扼要地重述为如下的形式定义：

定义. 所謂一个随机变量 $\mathbf{X}$，卽是定义在一个給定的样本空

間上的函数，也就是对每一个样本点給予一个实数与其对应．方程組(1.1)定义 $\mathbf{X}$ 的(概率)分布．如果两个随机变量 $\mathbf{X}$ 和 $\mathbf{Y}$ 定义在同一个样本空間中，則其联合分布由(1.3)所給出，它給出了 $\mathbf{X}$ 和 $\mathbf{Y}$ 取一切組合(x_j, y_k) 的概率．这个概念可以推广到定义在同一个样本空間中的任意有限个随机变量 $\mathbf{X}, \mathbf{Y}, \cdots, \mathbf{W}$ 的情形．如果对任意一組值$(x, y, \cdots, w)$都有

$$\mathbf{P}\{\mathbf{X}=x, \mathbf{Y}=y, \cdots, \mathbf{W}=w\}= \\ =\mathbf{P}\{\mathbf{X}=x\}\mathbf{P}\{\mathbf{Y}=y\}\cdots\mathbf{P}\{\mathbf{W}=w\}, \qquad (1.13)$$

則說随机变量 $\mathbf{X}, \mathbf{Y}, \cdots, \mathbf{W}$ 是相互独立的．

在第五章第 4 节中我們曾經定义了 n 次相互独立試驗的样本空間．把这个定义与(1.13)比較，我們看到，如果 $\mathbf{X}_k$ 只依賴于第 k 次試驗的結果，則随机变量 $\mathbf{X}_1, \cdots, \mathbf{X}_n$ 是相互独立的．更一般地，如果随机变量 U 只依賴于前 k 次試驗的結果，而另一随机变量 V 只依賴于后面 $n-k$ 次試驗的結果，則 U, V 是相互独立的(参看問題 39)．

我們可以把随机变量想象为样本空間中的点子的一种标記．这种手續在骰子的例子中是很熟悉的，在骰子上每一面都用数目来标明，而且我們把数目說成是单独試驗的可能的結果．在传統的数学术語里，我們說随机变量是由原始样本空間到新空間上的映象，这个新空間的点子是 $x_1, x_2, \cdots$，因此有：

当$\{f(x_j)\}$滿足条件(1.2)时，不再涉及原来的样本空間，我們就可以說一个随机变量 $\mathbf{X}$ 以概率 $f(x_1), f(x_2), \cdots$ 来取值 $x_1, x_2, \cdots$；新的样本空間是由样本点 $x_1, x_2, \cdots$ 构成的．指定一个概率分布等价于指定一个点为实数的样本空間．談及两个相互独立的随机变量 $\mathbf{X}, \mathbf{Y}$ 具有分布$\{f(x_j)\}, \{g(y_k)\}$等价于談及一个以数对(x_j, y_k)为样本点的样本空間，且其概率由 $\mathbf{P}\{(x_j, y_k)\}=f(x_j)g(y_k)$ 来确定．同样，对一組 n 个随机变量$(\mathbf{X}, \mathbf{Y}, \cdots, \mathbf{W})$所对应的样本空間，我們可以取 n 維空間中的点$(x, y, \cdots, w)$所构成的点集，而其中的概率由联合分布所給出．如果联合分布由(1.13)所給出，則說随机变量是相互独立的．

例. (e) 具有可变概率的伯努利試驗. 考虑 n 次相互独立的試驗,每一次試驗都只有两个結果: S 和 F. 第 k 次試驗出現 S 的概率为 p_k, 出現 F 的概率为 $q_k = 1 - p_k$. 如果 $p_k = p$, 則这个方案就变为伯努利試驗. 最簡单的描述它的方法是把 S 和 F 分別賦以值 1 和 0. 因此,这个模型完全可以用下面的語言来描述: 我們有 n 个相互独立的随机变量序列 $\mathbf{X}_k$, $\mathbf{X}_k$ 的分布为 $\mathbf{P}\{\mathbf{X}_k = 1\} = p_k$, $\mathbf{P}\{\mathbf{X}_k = 0\} = q_k$. 这个方案可以了解为"普阿松試驗"[参看例(5.b)和第十一章(6.b)].

显見,同一个分布可以联系不同的样本空間. 如果我們說随机变量 $\mathbf{X}$ 各以 1/2 的概率取值 1 和 0,則我們就默認了一个由0,1两点所构成的样本空間. 然而,我們也可以用这样的約定来定义随机变量 $\mathbf{X}$: 把一个錢币扔十次,若第十次出現正面,則 $\mathbf{X}$ 取 0;出現反面,則取 1. 这时 $\mathbf{X}$ 就定义在由所有的序列 $(HHT\cdots)$ 所构成的样本空間上,这个样本空間有 2^{10} 个样本点.

原則上,我們能够把概率論限制在由随机变量的概率分布概念所确定的样本空間上. 这样作就避免涉及抽象的样本空間,也避免涉及到"試驗","实驗的結果"等术語.把概率論化为随机变量的理論立刻导致分析知識可以用上而且在各方面簡化了理論. 然而,它也有美中不足之处,那就是使得概率背景晦澀不明. 随机变量的概念仍然容易含糊地当作"某些以不同的概率取不同的值的东西",但是随机变量本身是通常的函数,而函数这个概念并不是概率論所特有的.

例. (f) 令 $\mathbf{X}$ 为一个随机变量,其可能取的值为 $x_1, x_2, \cdots$, 且对应的概率为 $f(x_1), f(x_2), \cdots$. 为了帮助讀者想象,可以构造一个概念性的实驗来导出 $\mathbf{X}$. 例如把輪盘賭具的輪盘分成弧 $l_1, l_2, \cdots$, 其长度为 $f(x_1): f(x_2): \cdots$. 如果輪盘賭在弧 l_j 中的某一点停下来,則我們可假想某一个賭徒获得一笔款項 x_j(正的或者負的). 于是 $\mathbf{X}$ 就是那个賭徒所获的利益. 在 n 次試驗中所获之利益假定为 n 个具有共同分布 $\{f(x_j)\}$ 的相互独立的随机变量. 如討每一个組合(x_j, y_k)都有一段弧与之对应,且这时想象两个賭

徒各获利 x_j, y_k, 便得到具有給定的联合分布$\{p(x_j, y_k)\}$的两个随机变量.

若 $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \cdots$ 为定义在同一个样本空間上的随机变量，則任一函数 $F(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \cdots)$仍然是一个随机变量. 其分布可以从 $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \cdots$的联合分布得出,即簡单地把 $F(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \cdots)$这些取同一值时的所有的組合$(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \cdots)$的概率加起来.

例. (g) 在表 2 所表明的例子中,和 $\mathbf{X}_1+\mathbf{X}_2$ 是一个随机变量,其可能取的值为 0, 1, 2, 3,且其对应的概率为 q, $6q$, $12q$, $8q$ $\left(\text{其中 } q = \frac{1}{27}\right)$. 积 $\mathbf{X}_1\mathbf{X}_2$ 是另一随机变量，其可能取的值是 0, 1, 2, 其相应的概率为 $15q$, $6q$, $6q$.

2. 期 望 值

为了得到合理的簡化，常常需要用一些“特征值”来概要地描述概率分布. 在以前的关于等待时間的問題中曾經涉及过的中位数就是一个例子. 分布(1.1)的中位数 x_m 滿足条件: $\mathbf{P}\{\mathbf{X} \leqslant x_m\} \leqslant \frac{1}{2}$ 和 $\mathbf{P}\{\mathbf{X} \geqslant x_m\} \leqslant \frac{1}{2}$ 的数値 x_m. 換句話說,中位数 x_m 是这样一个数,它使得 $\mathbf{X}$ 超过它或者少于它的概率都接近 1/2.

然而,在这些特征值中，期望值或者均值是非常重要的一个. 一則对于它易于作分析处理,二則它有一个好的性質,即抽样的稳定性而使得統計学家們喜欢用它. 它的定义来自习惯上的平均概念. 假定在由家庭所构成的总体中，有 k 个小孩的家庭有 n_k 个，于是家庭的总数为 $n = n_0 + n_1 + n_2 + \cdots$, 而小孩的总数为 $m = n_1 + 2n_2 + 3n_3 + \cdots$. 故每个家庭的小孩平均値为 m/n. 概率与频率之間的类似性启发我們給出下面的定义:

定义. 令 $\mathbf{X}$ 为一个随机变量，其可能取的值为 $x_1, x_2, \cdots$, 且其对应的概率为 $f(x_1), f(x_2), \cdots$, 若下列級数

$$\mathbf{E}(\mathbf{X}) = \sum x_k f(x_k) \tag{2.1}$$

絕对收斂,則定义(2.1)右边的級数为 $\mathbf{X}$ 的期望値或者均值. 这时我們說 $\mathbf{X}$ 有有限的期望值. 如果 $\sum |x_k| f(x_k)$ 发散,則說 $\mathbf{X}$ 沒有

有限的期望值。

不用説，最常用的随机变量都是有有限的期望値的；否则这个概念就沒有实际意义了．然而与物理学中一些重要的复現問題相联系的随机变量就沒有有限的期望値．“均値”、“平均数”、“数学期望”、“期望値”这些术語都是同义詞．我們有时也説分布的均値来代替随机变量的均値．在数学和統計学中，一般采用 $\mathbf{E}(\mathbf{X})$ 这一符号来表示 $\mathbf{X}$ 之期望値，而在物理学中則通常用 $\overline{\mathbf{X}}, \langle\mathbf{X}\rangle, \langle\mathbf{X}\rangle_{Av}$ 来代替符号 $\mathbf{E}(\mathbf{X})$。

我們希望計算例如説 $\mathbf{X}^2$ 的期望値．这个函数是一个新的随机变量，其所取的可能値为 x_k^2；一般地説 $\mathbf{X}^2 = x_k^2$ 的概率不是 $f(x_k)$ 而是 $f(x_k) + f(-x_k)$．而且定义 $\mathbf{E}(\mathbf{X}^2)$ 为 $\sum' x_k^2\{f(x_k) + f(-x_k)\}$，求和上一撇表示只对 $x_k \geqslant 0$ 时求和．显見

$$\mathbf{E}(\mathbf{X}^2) = \sum x_k^2 f(x_k), \tag{2.2}$$

如果右边的級数收斂的話。

更一般地，我們用同样的方法可得出：

定理 1. 任一函数 $\phi(x)$，定义一个新的随机变量 $\phi(\mathbf{X})$．如果 $\phi(\mathbf{X})$具有有限期望値，則

$$\mathbf{E}(\phi(\mathbf{X})) = \sum \phi(x_k) f(x_k), \tag{2.3}$$

此处的級数絕对收斂当且仅当 $\mathbf{E}(\phi(\mathbf{X}))$ 存在．对任一常数 a，我們有 $\mathbf{E}(a\mathbf{X}) = a\mathbf{E}(\mathbf{X})$。

如果几个随机变量 $\mathbf{X}_1, \cdots, \mathbf{X}_n$ 定义在同一个样本空間上，則它們的和 $\mathbf{X}_1 + \cdots + \mathbf{X}_n$ 是一个新的随机变量．它的可能値与对应的概率可以从 $\mathbf{X}_\nu$ 的联合分布找出，从而 $\mathbf{E}(\mathbf{X}_1 + \cdots + \mathbf{X}_n)$ 可以算出．下面的重要定理提供了一个簡便的計算方法：

定理 2. 如果 $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ 都是具有有限期望値的随机变量，則它們的和的期望値存在，且其和的期望値就等于期望値的和：

$$\mathbf{E}(\mathbf{X}_1 + \cdots + \mathbf{X}_n) = \mathbf{E}(\mathbf{X}_1) + \cdots + \mathbf{E}(\mathbf{X}_n). \tag{2.4}$$

証．只需对两个随机变量 $\mathbf{X}, \mathbf{Y}$ 的情形来証明(2.4)就够了．应用(1.3)的符号，我們能够写出

$$\mathbf{E}(\mathbf{X})+\mathbf{E}(\mathbf{Y})=\sum_{j,k} x_j p(x_j, y_k)+\sum_{j,k} y_k p(x_j, y_k), \qquad (2.5)$$

这个求和法跑遍一切可能的值 x_j, y_k（它們不需要完全不同）. 由于 (2.5) 中之两級数絕对收斂，故其和可重排次序，从而得出 $\sum_{j,k}(x_j+y_k)p(x_j, y_k)$. 由期望值的定义得知，这就是 $\mathbf{X}+\mathbf{Y}$ 之期望值. 于是，我們完成了定理的証明.

显然，对于随机变量的积，是沒有与上述定理对应的一般定理. 例如 $\mathbf{E}(\mathbf{X}^2)$ 一般地并不等于 $\mathbf{E}(\mathbf{X})^2$. 如 $\mathbf{X}$ 是擲一顆均匀的骰子所得的点数，则 $\mathbf{E}(\mathbf{X})=\frac{7}{2}$，但 $\mathbf{E}(\mathbf{X}^2)=(1+4+9+16+25+36)/6=91/6$. 然而，对相互独立的随机变量来說，簡单的乘法規則是成立的，这就是下面的定理:

定理 3. 如果 $\mathbf{X}, \mathbf{Y}$ 是具有有限期望的相互独立的随机变量，则它們的积也是具有有限期望值的随机变量，且

$$\mathbf{E}(\mathbf{XY})=\mathbf{E}(\mathbf{X})\mathbf{E}(\mathbf{Y}). \qquad (2.6)$$

証. 为了計算 $\mathbf{E}(\mathbf{XY})$，我們将对每一个可能值 $x_j y_k$ 乘上其对应的概率. 前面已經指出，在定义(2.1)中 x_k 之值不一定完全不同. 因此

$$\mathbf{E}(\mathbf{XY})=\sum_{j,k} x_j y_k f(x_j) g(y_k)=$$

$$=\left\{\sum_j x_j f(x_j)\right\}\left\{\sum_k y_k g(y_k)\right\}. \qquad (2.7)$$

上面关于級数重排次序是可以允許的，因为这些級数为絕对收斂. 这就証明了我們的定理. 由归納法可推出: 对任意有限个相互独立的随机变量，类似的乘法規則还是成立的.

为了方便起見，我們給出条件概率分布的数学期望的概念. 如果 $\mathbf{X}$ 和 $\mathbf{Y}$ 是服从联合分布(1.3)的两个随机变量，給定 $\mathbf{X}$ 后 $\mathbf{Y}$ 的条件期望 $\mathbf{E}(\mathbf{Y}|\mathbf{X})$ 由下式:

$$\sum_k y_k \mathbf{P}\{\mathbf{Y}=y_k|\mathbf{X}=x_j\}=\frac{\sum_k y_k p(x_j, y_k)}{f(x_j)} \qquad (2.8)$$

定义，如果級数絕对收斂，而且对一切 j 有 $f(x_j)>0$.

3. 例子及应用

(a) 二項分布． 令 $\mathbf{S}_n$ 是在 n 次成功概率为 p 的伯努利試験中的成功次数． 我們知道 $\mathbf{S}_n$ 具有二項分布 $\{b(k; n, p)\}$． 因此 $\mathbf{E}(\mathbf{S}_n) = \sum kb(k; n, p) = np\sum b(k-1; n-1, p)$，而最后一个級数包含了对应于 $n-1$ 的二項分布的一切項，故其和为 1． 因此二項分布的均值为

$$\mathbf{E}(\mathbf{S}_n) = np. \tag{3.1}$$

不需要直接計算而采用一种常用的方便方法仍可得出同样的結果． 令 $\mathbf{X}_k$ 为第 k 次試験中成功的次数，則这个随机变量只取值 0，1，且其对应的概率为 p, q． 因此 $\mathbf{E}(\mathbf{X}_k) = 0\cdot q + 1\cdot p = p$． 又因

$$\mathbf{S}_n = \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_n, \tag{3.2}$$

故从(2.4)便可直接得出(3.1)．

(b) 普阿松分布． 如果 $\mathbf{X}$ 具有普阿松分布 $p(k;\lambda) = e^{-\lambda}\lambda^k/k!$（其中 $k = 0, 1, 2, \cdots$）． 則 $\mathbf{E}(\mathbf{X}) = \sum kp(k;\lambda) = \lambda\sum p(k-1;\lambda)$，而最后一个級数包含了普阿松分布的一切項，故其和为 1． 因此普阿松分布之均值为 λ．

(c) 負二項分布． 令 $\mathbf{X}$ 为具有几何分布 $\mathbf{P}\{\mathbf{X} = k\} = q^kp$（$k = 0, 1, 2, \cdots$）的随机变量． 于是 $\mathbf{E}(\mathbf{X}) = qp(1 + 2q + 3q^2 + \cdots)$． 右边是一个几何級数的导数，所以 $\mathbf{E}(\mathbf{X}) = qp(1-q)^{-2} = q/p$． 在第六章第 8 节我們曾經看到 $\mathbf{X}$ 可以解释为一个伯努利試験序列中第一次成功以前失敗的次数． 更一般地，我們研究过对应于直到第 n 次成功出現为止的伯努利試験的样本空間． 当 $r < n$ 时，令 $\mathbf{X}_1 = \mathbf{X}$，令 $\mathbf{X}_r$ 为从第 $r-1$ 次成功之后到第 r 成功之前的失敗次数． 于是每一个 $\mathbf{X}_\nu$ 都具有几何分布 $\{q^kp\}$，而且 $\mathbf{E}(\mathbf{X}_\nu) = q/p$． 和 $\mathbf{Y}_r = \mathbf{X}_1 + \cdots + \mathbf{X}_r$ 是第 r 次成功以前失敗的次数． 換句話說，$\mathbf{Y}_r$ 是一个随机变量，它的分布是如第六章(8.1)或等价的(8.2)所定义的負二項分布． 由此推出，这个負二項分布的均值为 rq/p． 这也可以用直接的計算来驗証． 显然，从第六章(8.2)推

出 $kf(k; r, p) = rp^{-1}qf(k-1; r+1, p)$，而这个分布的諸項之和为 1. 这个直接的計算有一个好处：它也可以应用到非整数 r 上去. 另一方面，第一种推导不需要知道 $\mathbf{X}_1 + \cdots + \mathbf{X}_r$ 的分布的显明形式就可以得出結果.

（d）抽样的等待时間. 从具有 n 个不同元素的总体中，作有放回的抽样. 由于重迭性，大小为 r 的随机样本中所包含的不同的元数一般地小于 r 个. 当样本的大小增加时，新元素进入样本就会愈来愈小. 在这个試驗中有一个很有趣的随机变量，那就是得到 r 个不同元素时的样本大小 $\mathbf{S}_r$.（例如，考虑 $N = 365$ 个可能的生日，这里 $\mathbf{S}_r$ 就表示：当样本中包含了 r 个不同生日的人时，所抽出来的人的数目. 类似的解釋可以用到球放入盒中的例子. 贈券或其他玩物的收集者对我們的問題特別感兴趣，如果他得贈券的情况可以比作随机抽样的話[1].）

第一个元素在第一次抽取时进入样本，从第二次抽取一直到另一新元素进入样本时所抽取的次数是一个随机变量 $\mathbf{X}_1$，一般地令 $\mathbf{X}_r$ 为从第 r 个不同元素进入样本后到第 $r+1$ 个不同的元素进入样本所需之抽取次数. 于是 $\mathbf{S}_r = 1 + \mathbf{X}_1 + \cdots + \mathbf{X}_{r-1}$ 是第 r 个新元素进入样本时的样本大小. 当样本包含了 k 个不同的元素时，每次抽取中获得新元素的概率为 $p = (N-k)/N$. 到达而且包含抽出一个新元素的次数 $\mathbf{X}_k$ 等于概率为 $p = (N-k)/N$ 的伯努利試驗第一次成功前的失敗次数加上 1. 因此，$\mathbf{E}(\mathbf{X}_k) = 1 + \frac{q}{p} = N/(N-k)$，而且从加法定理(2.4)得出

$$\mathbf{E}(\mathbf{S}_r) = N\left\{\frac{1}{N} + \frac{1}{N-1} + \frac{1}{N-2} + \cdots + \frac{1}{N-r+1}\right\}. \tag{3.3}$$

对 $r = N$，我們得到把总体所有的元素都抽到的期望值. 如 $N = 10$，则 $\mathbf{E}(\mathbf{S}_{10}) = 29.29\cdots$，$\mathbf{E}(\mathbf{S}_5) = 6.46\cdots$. 这就意味着，

1）参看[66]. 波利亚用不同的方法处理了一个稍为更一般的問題. 討論贈券收集問題的文献是很多的.

我們可以期望大約以 6 次到 7 次的抽取把总体中的一半元素抽出来，而剩下那一半則需要 23 次以上抽取才能抽出来．对相当大的 N，(3.3)有一个良好的逼近：

$$\mathbf{E}(\mathbf{S}_r) \approx N \log \frac{N}{N-r+1}. \tag{3.4}$$

特別，对任一分数 $\alpha<1$，当 N 充分大时，要求样本中包含的不同元素为总体的 α 份时所需之抽取次数的期望值近似地为 $N \log \frac{1}{1-\alpha}$；要求样本中包含全部 N 个元素所需的抽取次数的期望值近似地为 $N \log N$．注意，在这里我們又沒有用分布就得出了上述的結果．

(e) 一个估計問題． 一个碗中裝有 N 个标有号数 1 到 N 的球．令 $\mathbf{X}$ 为采用有放回的随机抽样时，n 次抽取中所抽出来的最大的号数．事件 $\mathbf{X} \leqslant k$ 意味着所抽出来的 n 个号数中每一个都小于等于 k，因此 $\mathbf{P}\{\mathbf{X} \leqslant k\} = \left(\frac{k}{N}\right)^n$．所以 $\mathbf{X}$ 的概率分布由

$$\begin{aligned} p_k = \mathbf{P}\{\mathbf{X}=k\} &= \mathbf{P}\{\mathbf{X} \leqslant k\} - \mathbf{P}\{\mathbf{X} \leqslant k-1\} = \\ &= \{k^n - (k-1)^n\} N^{-n} \end{aligned} \tag{3.5}$$

所給出．故有

$$\begin{aligned} \mathbf{E}(\mathbf{X}) = \sum_{k=1}^{N} k p_k &= N^{-n} \sum_{k=1}^{n} \{k^{n+1} - (k-1)^{n+1} - (k-1)^n\} = \\ &= N^{-n} \left\{ N^{n+1} - \sum_{k=1}^{N} (k-1)^n \right\}. \end{aligned} \tag{3.6}$$

对于相当大的 N，最后一个和近似地为四条曲綫 $y = x^n$，$x = 0$，$x = N$ $y = 0$ 所围成之面积．也就是說，它等于 $N^{n+1}/(n+1)$．由此推出，对于相当大的 N，有

$$\mathbf{E}(\mathbf{X}) \approx \frac{n}{n+1} N. \tag{3.7}$$

如果一个城市有 $N = 1000$ 輛汽車．观察一个 $n = 10$ 的样本，由(3.7)知其执照牌上的最大的号碼（假定随机性）的期望数大約是 910（中位数为 934)．应用統計学家利用由样本中所观察到的最大値来估計未知的眞实数目 N．在第二次大战期間，曾用此法

估計敌人的生产[参看問題 8—11].

(f) 巴拿赫火柴盒問題. 在第六章第 8 节中我們求出,分布

$$u_r = \binom{2N - r}{N} \frac{1}{2^{2N-r}} \tag{3.8}$$

是第一次发現空盒时而另一个盒中所剩下的火柴的根数 $\mathbf{X}$ 的分布. 我們不能用直接的方法去計算期望值 $\mathbf{E}(\mathbf{X}) = \mu$, 但是下面的間接的办法在很多情況下都是可以应用的. 利用 u_r 之和为 1 (这是不容易驗証的)这一事实,我們求得

$$N - \mu = \sum_{r=0}^{N-1} (N - r) u_r = \sum_{r=0}^{N-1} (N - r) \binom{2N - r}{N - r} \frac{1}{2^{2N-r}}. \tag{3.9}$$

对二項式系数加以簡单运算,得知最后的和变为

$$\sum_{r=0}^{N-1} (2N - r) \binom{2N - r - 1}{N - r - 1} \frac{1}{2^{2N-r}} =$$

$$= \frac{2N + 1}{2} \sum_{r=0}^{N-1} u_{r+1} - \frac{1}{2} \sum_{r=0}^{N-1} (r + 1) u_{r+1}. \tag{3.10}$$

最后的和与定义 $\mu = \mathbf{E}(\mathbf{X})$ 的和是一样的. 第一个和包括除 u_0 以外一切其它的 u_r, 因此第一个和为 $1 - u_0$. 故由(3.9)与(3.10)得出

$$N - \mu = \frac{2N + 1}{2} (1 - u_0) - \frac{\mu}{2} \tag{3.11}$$

或

$$\mu = (2N + 1) u_0 - 1 = \frac{2N + 1}{2^{2N}} \binom{2N}{N} - 1. \tag{3.12}$$

应用史特令公式,我們发現

$$\mu \approx 2 \left(\frac{N}{\pi}\right)^{1/2} - 1. \tag{3.13}$$

特别,在第六章表 8 的分布中, $N = 50$ 这时 $\mu = 7.04\cdots$, 而中位数为 6.

4. 方　　差

令 $\mathbf{X}$ 为一个具有分布 $\{f(x_j)\}$ 的随机变量, 且令 $r \geq 0$ 为一

个整数．如果随机变量 $\mathbf{X}^r$ 的期望值存在，即

$$\mathbf{E}(\mathbf{X}^r) = \sum x_j^r f(x_j) \tag{4.1}$$

存在的話，则称它为 $\mathbf{X}$ 的 r 阶矩． 如果級数(4.1)并不絕对收斂，则說 $\mathbf{X}$ 的 r 阶矩不存在．因为 $|\mathbf{X}|^{r-1} \leqslant |\mathbf{X}|^r + 1$，故 r 阶矩存在时，$r-1$ 阶矩也存在，从而前面的各阶矩都存在．

在一般理論中，矩有重要的意义，然而在现在这一卷书里，我們仅仅用到二阶矩．如果二阶矩存在，当然均值

$$\mu = \mathbf{E}(\mathbf{X}) \tag{4.2}$$

也存在．于是，我們很自然地引进随机变量与其均值之差 $\mathbf{X}-\mu$ 来代替随机变量 $\mathbf{X}$．因为$(x-\mu)^2 \leqslant 2(x^2+\mu^2)$，故 $\mathbf{E}(\mathbf{X}^2)$ 存在时 $\mathbf{X}-\mu$ 的二阶矩也存在．我們有

$$\mathbf{E}((\mathbf{X}-\mu)^2) = \sum_j (x_j^2 - 2\mu x_j + \mu^2) f(x_j). \tag{4.3}$$

把右边分为三个单独的和，于是得知它等于 $\mathbf{E}(\mathbf{X}^2) - 2\mu\mathbf{E}(\mathbf{X}) + \mu^2 = \mathbf{E}(\mathbf{X}^2) - \mu^2$．

定义． 設 $\mathbf{X}$ 为具有二阶矩 $\mathbf{E}(\mathbf{X}^2)$ 的随机变量． 且令 $\mu = \mathbf{E}(\mathbf{X})$ 为其均值．我們定义数目

$$\mathrm{Var}(\mathbf{X}) = \mathbf{E}((\mathbf{X}-\mu)^2) = \mathbf{E}(\mathbf{X}^2) - \mu^2 \tag{4.4}$$

为 $\mathbf{X}$ 的方差．其正的平方根(或者 0)叫作 $\mathbf{X}$ 的标准差．

为簡单計，我們常常說一个分布的方差，而不說及随机变量．“离差”是現在通常采用的术語“方差”的同义詞．

例． (a) 若 $\mathbf{X}$ 各以 1/2 的概率取值 $\pm c$，則 $\mathrm{Var}(\mathbf{X}) = c^2$．

(b) 若 $\mathbf{X}$ 为一个均匀骰子上面所刻的点数，則 $\mathrm{Var}(\mathbf{X}) = \frac{1}{6}(1^2+2^2+\cdots+6^2) - (7/2)^2 = 35/12$．

(c) 对于普阿松分布 $p(k;\lambda)$来說，其均值为 λ [参看(3.b)]，因此其方差为 $\sum k^2 p(k;\lambda) - \lambda^2 = \lambda\sum kp(k-1;\lambda) - \lambda^2 = \lambda\sum(k-1)p(k-1;\lambda) + \lambda\sum p(k-1;\lambda) - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$．在这个例子中，均值与方差相等．

(d) 对二項分布来說[参看(3.a)]，經过类似的計算得出其方差为

$$\sum k^2 b(k;n,p)-(np)^2 = np\sum k b(k-1;n-1,p)-(np)^2 =$$
$$= np\{(n-1)p+1\}-(np)^2 = npq.$$

方差这一概念的用处是逐漸地会显現出来的，特别是与和极限定理(第十章)联系起来的时候．在这里我們会看到：方差是松散度的粗略度量．事实上，如果 $\mathrm{Var}(\mathbf{X})=\sum(x_j-\mu)^2 f(x_j)$ 較小，則和中每一項都較小，从而对应于"使 $|x_j-\mu|$ 較大的" x_j 的概率 $f(x_j)$ 較小．換言之，在方差小的时候，$\mathbf{X}$ 与 μ 之大偏差是不概然的．反过来，方差大时，則 $\mathbf{X}$ 的可能值不会完全落在其均值附近．

讀者們可以用下面的力学解释来帮助了解．假定一个单位的質量分布在 x 軸上，使得在点 x_j 集中了質量 $f(x_j)$，于是均值 μ 就是重心的横坐标，方差就是轉动慣量．显然，不同的質量分布可以有相同的重心和相同的轉动慣量，然而如所周知，很多重要的力学性質可以用这两个量来描述．

如果 $\mathbf{X}$ 代表一个可以測量的量，例如长度或者温度，則其数值依賴于原点及測量单位．而原点与单位的改变就意味着把 $\mathbf{X}$ 变化到一个新的变量 $a\mathbf{X}+b$，此处 a 和 b 都是常数．显然，$\mathrm{Var}(\mathbf{X}+b)=\mathrm{Var}(\mathbf{X})$，因而

$$\mathrm{Var}(a\mathbf{X}+b)=a^2\mathrm{Var}(\mathbf{X}). \tag{4.5}$$

原点及測量单位的选取有很大的任意性．不过通常选取均值作原点，标准差作測量的单位則是較方便的．我們在第七章中已經这样作了，当时我們引进了正則化成功次数 $\mathbf{S}_n^*=(\mathbf{S}_n-np)/(npq)^{1/2}$．一般地，如果 $\mathbf{X}$ 有均值 μ 与方差 $\sigma^2\ (\sigma>0)$，則 $\mathbf{X}-\mu$ 的均值为 0，方差为 σ^2，因此随机变量

$$\mathbf{X}^*=\frac{\mathbf{X}-\mu}{\sigma} \tag{4.6}$$

之均值为 0 方差为 1，它叫作 $\mathbf{X}$ 的正則化随机变量．在物理学家的語言中，从 $\mathbf{X}$ 变到 $\mathbf{X}^*$ 解释为引进无量綱的量．

5. 协方差；和的方差

設 $\mathbf{X}$ 与 $\mathbf{Y}$ 为同一样本空間中的两个随机变量，則 $\mathbf{X}+\mathbf{Y}$ 与

$\mathbf{XY}$ 仍然是随机变量. 它們的分布可以从 $\mathbf{X}$ 和 $\mathbf{Y}$ 的联合分布經过一些簡单的运算而得出. 我們現在的目的是計算 $\mathrm{Var}(\mathbf{X}+\mathbf{Y})$. 为此我們引进协方差的概念. 它将在第 8 节里詳細地进行分析. 如果 $\mathbf{X}$ 与 $\mathbf{Y}$ 的联合分布为$\{p(x_j, y_k)\}$,則 $\mathbf{XY}$ 的期望值由

$$\mathbf{E}(\mathbf{XY}) = \sum x_j y_k p(x_j, y_k) \tag{5.1}$$

所給出,当然,假定右边的級数是絕对收斂的. 由于 $|x_j y_k| \leqslant (x_j^2 + y_k^2)/2$, 故当 $\mathbf{E}(\mathbf{X}^2)$, $\mathbf{E}(\mathbf{Y}^2)$ 存在时, $\mathbf{E}(\mathbf{XY})$也存在. 这时,期望値

$$\mu_x = \mathbf{E}(\mathbf{X}), \quad \mu_y = \mathbf{E}(\mathbf{Y}) \tag{5.2}$$

当然也存在. 而且 $\mathbf{X} - \mu_x$, $\mathbf{Y} - \mu_y$ 之均値都是 0. 至于它們的积,由第 2 节的加法規則,有

$$\begin{aligned}\mathbf{E}((\mathbf{X} - \mu_x)(\mathbf{Y} - \mu_y)) &= \mathbf{E}(\mathbf{XY}) - \mu_x\mathbf{E}(\mathbf{Y}) - \\ &- \mu_y\mathbf{E}(\mathbf{X}) + \mu_x\mu_y = \mathbf{E}(\mathbf{XY}) - \mu_x\mu_y.\end{aligned} \tag{5.3}$$

定义. 我們定义

$$\mathrm{Cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{E}((\mathbf{X}-\mu_x)-(\mathbf{Y}-\mu_y)) = \mathbf{E}(\mathbf{XY}) - \mu_x\mu_y \tag{5.4}$$

为 $\mathbf{X}$, $\mathbf{Y}$ 的协方差. 当 $\mathbf{X}$ 和 $\mathbf{Y}$ 有有限的方差时,这个定义是有意义的.

从第 2 节中我們知道,对相互独立的随机变量来說,有$\mathbf{E}(\mathbf{XY}) = \mathbf{E}(\mathbf{X})\mathbf{E}(\mathbf{Y})$,因此由(5.4),有

定理 1. 如果 $\mathbf{X}$, $\mathbf{Y}$ 相互独立,則 $\mathrm{Cov}(\mathbf{X}, \mathbf{Y}) = 0$.

注意: 其逆不眞. 例如,閱看表 1 便知,那两个随机变量是不独立的,但是它們的协方差为 0. 在第 8 节我們还要回过来討論这一点. 下面的定理是很重要的,对独立的随机变量来說,加法規則(5.6)仍然成立.

定理 2. 如果 $\mathbf{X}_1, \cdots, \mathbf{X}_n$ 为具有有限方差 $\sigma_1^2, \cdots, \sigma_n^2$ 的 n 个随机变量,且令 $\mathbf{S}_n = \mathbf{X}_1 + \cdots + \mathbf{X}_n$, 則

$$\mathrm{Var}(\mathbf{S}_n) = \sum_{k=1}^{n} \sigma_k^2 + 2\sum_{j,k} \mathrm{Cov}(\mathbf{X}_j, \mathbf{X}_k), \tag{5.5}$$

此处最后一个求和包括了$\binom{n}{2}$对$(\mathbf{X}_j, \mathbf{X}_k)$而 $j < k$.

特别，如果 $\mathbf{X}_j$ 相互独立，则給出了加法規則：

$$\operatorname{Var}(\mathbf{S}_n) = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2. \tag{5.6}$$

証. 令 $\mu_k = \mathbf{E}(\mathbf{X}_k)$; $m_n = \mu_1 + \cdots + \mu_n = \mathbf{E}(\mathbf{S}_n)$，则 $\mathbf{S}_n - m_n = \sum(\mathbf{X}_k - \mu_k)$ 且

$$(\mathbf{S}_n - m_n)^2 = \sum(\mathbf{X}_k - \mu_k)^2 + 2\sum(\mathbf{X}_j - \mu_j)(\mathbf{X}_k - \mu_k). \tag{5.7}$$

把(5.7)两边取期望值，然后应用加法規則即可得出(5.5). 至于(5.6)的証明，只需应用定理 1 即可.

例. (a) 二項分布 $\{b(k; n, p)\}$. 在例(3.a)中，$\mathbf{X}_k$ 是相互独立的. 我們有 $\mathbf{E}(\mathbf{X}_k^2) = 0^2 \cdot q + 1^2 \cdot p = p$，$\mathbf{E}(\mathbf{X}_k) = p$，因此，$\sigma_k^2 = p - p^2 = pq$. 且从(5.6)得知二項分布之方差为 npq. 在例(4.d)中，用直接計算的办法已經得出同样的結果.

(b) 具有可变概率的伯努利試驗. 令 $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ 为相互独立的随机变量，$\mathbf{X}_k$ 分别以概率 p_k 和 $q_k = 1 - p_k$ 取值 1 和 0. 于是 $\mathbf{E}(\mathbf{X}_k) = p_k$，$\operatorname{Var}(\mathbf{X}_k) = p_k - p_k^2 = p_k q_k$. 再令 $\mathbf{S}_n = \mathbf{X}_1 + \cdots + \mathbf{X}_n$，从(5.6)我們得到

$$\operatorname{Var}(\mathbf{S}_n) = \sum_{k=1}^{n} p_k q_k. \tag{5.8}$$

如例(1.e)一样，随机变量 $\mathbf{S}_n$ 可以解释为 n 次独立的試驗成功的次数，每一次試驗的結果或者成功或者失敗. 于是 $p = (p_1 + \cdots + p_n)/n$ 是其成功的平均概率，很自然地，我們要把現在的情形和具有固定成功概率 p 的伯努利試驗比較. 这样一种比較可以得出一种很好的結果. 我們可以把(5.8)写为 $\operatorname{Var}(\mathbf{S}_n) = np - \sum p_k^2$. 其次，容易看出(用初等的計算或者簡单的归納法)，在全部滿足 $\sum p_k = np$ 的組 $\{p_k\}$ 中，当所有的 p_k 都相等时，$\sum p_k^2$ 达到它的最小值. 由此推出，如果成功的平均概率 p 固定，则当 $p_1 = p_2 = \cdots = p_k = p$ 时，$\operatorname{Var}(\mathbf{S}_n)$ 达到最大值. 因此，我們得到了一个很好的結果：p_k 的可变性或者机遇的不一致性減小了其随机起伏性[1)]（用方差来衡量的）. 例如，一个城市中一年的火災可以考虑为一

1) 更強的結果參看[67].

个随机变量；对于一个給定的平均数，如果每一戶发生火災的概率都一样，則火災的变化性达到最大． 給定 n 个机器的某一质量平均 p，如果全部机器都一样，則产品最不一致．（把它应用到近代教育中去是明显的但是沒有什么用处．）

(c) 合牌． 一副具有 n 张标有号数的紙牌随机地排列，于是 $n!$ 种排列都是等概的．相合（牌在它們自然的位置）的数目是一个取值为 $0, 1, \cdots, n$ 的随机变量 $\mathbf{S}_n$，其概率分布已經在第四章中推导过．由它的分布我們可以得出 $\mathbf{S}_n$ 之均值与方差． 不过下面的方法較簡单而且很有启发性．

我們定义一个随机变量 $\mathbf{X}_k$，其所取的值为 1 或 0，当号碼为 k 的牌在第 k 个位置时 $\mathbf{X}_k$ 为 1，反之 $\mathbf{X}_k$ 为 0． 于是 $\mathbf{S}_n = \mathbf{X}_1 + \cdots + \mathbf{X}_n$．每一张牌都以概率 $1/n$ 出現在第 k 个位置．故 $\mathbf{P}\{\mathbf{X}_k = 1\} = \frac{1}{n}$ 而 $\mathbf{P}\{\mathbf{X}_k = 0\} = \frac{n-1}{n}$，因此 $\mathbf{E}(\mathbf{X}_k) = \frac{1}{n}$． 由此可知 $\mathbf{E}(\mathbf{S}_n) = 1$，亦即一副紙牌中相合数目的均值为 1． 为了計算其方差，我們首先計算 $\mathbf{X}_k$ 的方差 σ_k^2：

$$\sigma_k^2 = \frac{1}{n} - \left(\frac{1}{n}\right)^2 = \frac{n-1}{n^2}. \tag{5.9}$$

其次我們計算 $\mathbf{E}(\mathbf{X}_j \mathbf{X}_k)$．积 $\mathbf{X}_j\mathbf{X}_k$ 或者为 0 或者为 1，且只有当号碼为 j 的牌在第 j 个位置，号碼为 k 的牌在第 k 个位置时，$\mathbf{X}_j\mathbf{X}_k$ 才为 1．这个事件的概率为 $1/n(n-1)$．故

$$\mathbf{E}(\mathbf{X}_j\mathbf{X}_k) = \frac{1}{n(n-1)},$$

$$\mathrm{Cov}(\mathbf{X}_j, \mathbf{X}_k) = \frac{1}{n(n-1)} - \frac{1}{n^2} = \frac{1}{n^2(n-1)}. \tag{5.10}$$

故最后有

$$\mathrm{Var}(\mathbf{S}_n) = n\frac{n-1}{n^2} + 2\binom{n}{2}\frac{1}{n^2(n-1)} = 1. \tag{5.11}$$

从上面我們看到了相合数目的均值与方差都等于 1． 这个結果可以应用到第四章第 4 节中所討論的牌的猜測問題上去．那里我們考虑三种猜測的方法，其中之一就对应于合牌． 第二种可描述为

概率 $p=\frac{1}{n}$ 的 n 次伯努利試驗序列． 在这种情形下，猜对的期望数为 $np=1$，方差为 $npq=(n-1)/n$． 上述两种情形中期望数都是一样的．但是第一种方法却具有較大的方差，这就表示围繞着均值有較大的偶然起伏，从而游戏可能較为紧张些．（对于較复杂的紙牌来說这两种方差之間的差别稍为大一点．但并不怎么大．）对最后一个猜測模型来說，猜的人保持呼喚同一张牌；猜对次数永远是 1，而偶然性的起伏完全消失（方差为 0）．我們看到，呼喚的策略不会影响猜对的期望数．但是它对偶然性的起伏的大小却有一些影响．

（d）无放回的抽样． 假定一个总体由 b 个黑元素与 g 个綠元素組成，而且假定由其中抽取一个大小为 r 的随机样本（不可能重选）．样本中黑元素的个数 $\mathbf{S}_r$ 是一个随机变量，它具有超几何分布（第二章第 6 节），均值和方差可以由直接的計算而得，但是下面的方法是較好的．定义一个只取值 1 或者 0 的随机变量 $\mathbf{X}_k$，当样本中第 k 个元素是黑元素时 $\mathbf{X}_k$ 为 1（$k\leqslant r$），否則 $\mathbf{X}_k$ 为 0．由于对称性的原故，$\mathbf{X}_k=1$ 的概率为 $b/(b+g)$，从而

$$\mathbf{E}(\mathbf{X}_k)=\frac{b}{b+g},\quad \operatorname{Var}(\mathbf{X}_k)=\frac{bg}{(b+g)^2}. \tag{5.12}$$

其次，如果 $j\neq k$，则当样本中第 j 个和第 k 个元素都是黑时，$\mathbf{X}_j\mathbf{X}_k=1$，否則 $\mathbf{X}_j\mathbf{X}_k=0$．$\mathbf{X}_j\mathbf{X}_k=1$ 的概率为 $b(b-1)/(b+g)\times(b+g-1)$．因此

$$\begin{aligned}\mathbf{E}(\mathbf{X}_j\mathbf{X}_k)&=\frac{b(b-1)}{(b+g)(b+g-1)},\\ \operatorname{Cov}(\mathbf{X}_j\mathbf{X}_k)&=\frac{-bg}{(b+g)^2(b+g-1)}.\end{aligned} \tag{5.13}$$

所以

$$\mathbf{E}(\mathbf{S}_r)=\frac{rb}{b+g},\quad \operatorname{Var}(\mathbf{S}_r)=\frac{rbg}{(b+g)^2}\left\{1-\frac{r-1}{b+g-1}\right\}. \tag{5.14}$$

在有放回的抽样中，我們仍有同样的均值，但其方差稍为大一点，其方差为 $rbg/(b+g)^2$．

6. 切比謝夫不等式[1]

曾經指出过，小的方差表示与均值的大偏差是不概然的． 这一事实由切比謝夫不等式而变得更精确了．这个不等式是一个很有用而且很方便的工具．

定理． 令 $\mathbf{X}$ 为一个具有均值 $\mu = \mathbf{E}(\mathbf{X})$，方差 $\sigma^2 = \mathrm{Var}(\mathbf{X})$ 的随机变量．于是对于任一 $t > 0$，都有

$$\mathbf{P}\{|\mathbf{X} - \mu| \geqslant t\} \leqslant \frac{\sigma^2}{t^2}. \tag{6.1}$$

証． 方差是由（4.3）中的正項級数所定义的． 所以除去使 $|x_j - \mu| < t$ 的那些項以后，不会使級数的值增加．故

$$\sigma^2 \geqslant \sum{}^*(x_j - \mu)^2 f(x_j), \tag{6.2}$$

此处的星号（$*$）表示在求和时只对使 $|x_j - \mu| \geqslant t$ 的那些 j 来求和．故显然地有

$$\sum{}^*(x_j - \mu)^2 f(x_j) \geqslant t^2 \sum{}^* f(x_j) = t^2 \mathbf{P}\{|\mathbf{X} - \mu| \geqslant t\}. \tag{6.3}$$

这就証明了我們的定理．

切比謝夫不等式作为一个理論工具比作为估計的实际方法要恰当一些．其重要性在于它的应用普遍性．但是不能希望很普遍的命題会对一些个别情况給出深刻的結果．

例． (a)令$\mathbf{X}$为擲一顆均匀的骰子所得到的点数，则 $\mu = 7/2$，$\sigma^2 = 35/12$ [参看例(4.b)]． $\mathbf{X}$ 与 μ 之最大偏差为 $2.5 \approx 3\sigma/2$．$|\mathbf{X} - \mu|$ 大于这个偏差的概率为 0，然而切比謝夫不等式仅仅断定这个概率少于 0.47．

(b) 对二項分布$\{b(k; n, p)\}$来說，我們有 $\mu = np$，$\sigma^2 = npq$ [参看例(5.a)]．对于大的 n 我們知道

$$\mathbf{P}\{|\mathbf{S}_n - np| > x(npq)^{1/2}\} \approx 1 - \Phi(x) + \Phi(-x). \tag{6.4}$$

然而切比謝夫不等式只指出：左边少于 $1/x^2$；显然，这較之(6.4)来說是一个多么粗糙的估計．

---

1) P. L. Chebyshev (Чебышев, 1821—1894).

*7. 柯尔莫格洛夫不等式[68]

作为更精湛的方法的一个例子，我們証明：

令 $\mathbf{X}_1, \cdots, \mathbf{X}_n$ 为相互独立的，具有期望值 $\mu_k = \mathbf{E}(\mathbf{X}_k)$ 及方差 σ_k^2 的随机变量. 令

$$\mathbf{S}_k = \mathbf{X}_1 + \cdots + \mathbf{X}_k, \tag{7.1}$$

且

$$m_k = \mathbf{E}(\mathbf{S}_k) = \mu_1 + \cdots + \mu_k \tag{7.2}$$

$$s_k^2 = \mathrm{Var}(\mathbf{S}_k) = \sigma_1^2 + \cdots + \sigma_k^2.$$

对每一个 $t > 0$，下面 n 个不等式：

$$|\mathbf{S}_k - m_k| < ts_n \quad k = 1, 2, \cdots, n \tag{7.3}$$

同时成立的概率至少为 $1 - t^{-2}$.

当 $n = 1$ 时，这个定理化为切比謝夫不等式. 对 $n > 1$ 的情形，切比謝夫不等式只对每个单个的关系 $|\mathbf{S}_n - m_n| < ts_n$ 的概率給出了一样的界. 故柯尔莫格洛夫不等式是比較强的.

証. 我們要估計不等式(7.3)中有一个不成立的概率 x. 定理断言：$x \leqslant t^{-2}$.

定义 n 个随机变量 $\mathbf{Y}_k$ 如下. 如果

$$|\mathbf{S}_\nu - m_\nu| \geqslant ts_n, \tag{7.4}$$

且

$$|\mathbf{S}_k - m_k| < ts_n \text{ 对 } k = 1, 2, \cdots, \nu - 1, \tag{7.5}$$

则 $\mathbf{Y}_\nu = 1$；在其他一切样本点上，$\mathbf{Y}_\nu = 0$. 即是說，$\mathbf{Y}_\nu$ 在那些使得不等式(7.3)中第 ν 个不等式为首次不成立的样本点上为 1. 于是在任一特定的样本点上，諸 $\mathbf{Y}_k$ 中最多只有一个为 1，从而 $\mathbf{Y}_1 + \mathbf{Y}_2 + \cdots + \mathbf{Y}_n$ 只能取値 0 或者 1，当且仅当(7.3)中的不等式至少有一个不成立时它才为 1. 因此

$$x = \mathbf{P}\{\mathbf{Y}_1 + \cdots + \mathbf{Y}_n = 1\}. \tag{7.6}$$

因为 $\mathbf{Y}_1 + \cdots + \mathbf{Y}_n$ 为 0 或者为 1，所以 $\sum \mathbf{Y}_k \leqslant 1$. 两边乘以$(\mathbf{S}_n - m_n)^2$ 以后再取期望值，我們得到

$$\sum_{k=1}^{n} \mathbf{E}(\mathbf{Y}_k(\mathbf{S}_n - m_n)^2) \leqslant s_n^2. \tag{7.7}$$

为了計算左边每一項的值，我們令

$$\mathbf{U}_k=(\mathbf{S}_n-m_n)-(\mathbf{S}_k-m_k)=\sum_{\nu=k+1}^{n}(\mathbf{X}_\nu-\mu_\nu). \tag{7.8}$$

于是

$$\begin{aligned}&\mathbf{E}(\mathbf{Y}_k(\mathbf{S}_n-m_n)^2)\\&=\mathbf{E}(\mathbf{Y}_k(\mathbf{S}_k-m_k)^2)+2\mathbf{E}(\mathbf{Y}_k\mathbf{U}_k(\mathbf{S}_k-m_k))+\mathbf{E}(\mathbf{Y}_k\mathbf{U}_k^2). \end{aligned}\tag{7.9}$$

然而 $\mathbf{U}_k$ 只依賴于 $\mathbf{X}_{k+1},\cdots,\mathbf{X}_n$，而 $\mathbf{Y}_k$，$\mathbf{S}_k$，只依賴于 $\mathbf{X}_1,\cdots,\mathbf{X}_k$. 故 $\mathbf{U}_k$ 与 $\mathbf{Y}_k(\mathbf{S}_k-m_k)$ 相互独立. 因而 $\mathbf{E}(\mathbf{Y}_k\mathbf{U}_k(\mathbf{S}_k-m_k))=\mathbf{E}(\mathbf{Y}_k(\mathbf{S}_k-m_k))\mathbf{E}(\mathbf{U}_k)=0$ [因为 $\mathbf{E}(\mathbf{U}_k)=0$]. 故由(7.9)推出

$$\mathbf{E}(\mathbf{Y}_k(\mathbf{S}_n-m_n)^2)\geqslant\mathbf{E}(\mathbf{Y}_k(\mathbf{S}_k-m_k)^2). \tag{7.10}$$

但是只有当 $|\mathbf{S}_k-m_k|\geqslant ts_n$ 时，$\mathbf{Y}_k\neq0$. 故 $\mathbf{Y}_k(\mathbf{S}_k-m_k)^2\geqslant t^2s_n^2\mathbf{Y}_k$. 从而綜合(7.7)和(7.10)我們得出

$$s_n^2\geqslant t^2s_n^2\mathbf{E}(\mathbf{Y}_1+\cdots+\mathbf{Y}_n). \tag{7.11}$$

因为 $\mathbf{Y}_1+\cdots+\mathbf{Y}_n$ 或者为 0 或者为 1，故右边的那一个期望值等于(7.6)中所定义的概率 x. 因此 $xt^2\leqslant1$. 这就完成了我們的証明.

*8. 相关系数

令 $\mathbf{X}$, $\mathbf{Y}$ 为任意两个具有均值 μ_x, μ_y 和正方差 σ_x^2, σ_y^2 的随机变量. 我們引进如(4.6)中所定义的对应的正則化随机变量 $\mathbf{X}^*$, $\mathbf{Y}^*$. 它們的协方差称为 $\mathbf{X}$, $\mathbf{Y}$ 的相关系数，且以符号 $\rho(\mathbf{X},\mathbf{Y})$ 表示. 利用(5.4)，得到

$$\rho(\mathbf{X},\mathbf{Y})=\mathrm{Cov}(\mathbf{X}^*,\mathbf{Y}^*)=\frac{\mathrm{Cov}(\mathbf{X},\mathbf{Y})}{\sigma_x\sigma_y}. \tag{8.1}$$

显然，相关系数是不依賴于原点与測量单位的，也就是說，对于任意常数 a_1, a_2, b_1, b_2 (其中 $a_1>0$, $a_2>0$)都有 $\rho(a_1\mathbf{X}+b_1, a_2\mathbf{Y}+b_2)=\rho(\mathbf{X},\mathbf{Y})$.

使用相关系数，无非是給协方差的书写来一个花样罢了[1]. 很

1) 物理学家把相关系数定义作"无因次的协方差".

不幸，相关系数并沒有“相关”这个詞儿所暗示的涵义. 从第 5 节我們知道，当 $\mathbf{X}$、$\mathbf{Y}$ 相互独立时，$\rho(\mathbf{X}, \mathbf{Y}) = 0$. 但是应該知道，其逆命題是不成立的. 事实上，甚至当 $\mathbf{Y}$ 是 $\mathbf{X}$ 的函数时相关系数 $\rho(\mathbf{X}, \mathbf{Y})$也可能为 0.

例. (a) 令 $\mathbf{X}$ 各以 1/4 的概率取值 ± 1, ± 2, 且令 $\mathbf{Y} = \mathbf{X}^2$. 联合分布为 $p(-1, 1) = p(1, 1) = p(2, 4) = p(-2, 4) = \frac{1}{4}$. 虽然 $\mathbf{Y}$ 对 $\mathbf{X}$ 有直接的函数依賴性，可是由对称性的原故仍然有 $\rho(\mathbf{X}, \mathbf{Y}) = 0$.

(b) 令 $\mathbf{U}$, $\mathbf{V}$ 为相互独立的具有共同分布的随机变量，且令 $\mathbf{X} = \mathbf{U} + \mathbf{V}$, $\mathbf{Y} = \mathbf{U} - \mathbf{V}$. 于是 $\mathbf{E}(\mathbf{XY}) = \mathbf{E}(\mathbf{U}^2) - \mathbf{E}(\mathbf{V}^2)$, $\mathbf{E}(\mathbf{Y}) = 0$，故 $\mathrm{Cov}(\mathbf{X},\mathbf{Y}) = 0$，从而 $\rho(\mathbf{X},\mathbf{Y}) = 0$. 例如 $\mathbf{X}$, $\mathbf{Y}$ 可以分别为两顆骰子的点数之和与差，則 $\mathbf{X}$, $\mathbf{Y}$ 或者同时为奇数或者同时为偶数，所以 $\mathbf{X}$, $\mathbf{Y}$ 是非相互独立的.

由此推出，相关系数并不表示 $\mathbf{X}$ 和 $\mathbf{Y}$ 之間的依賴性的一般度量. 然而 $\rho(\mathbf{X}, \mathbf{Y})$ 与 $\mathbf{X}$ 和 $\mathbf{Y}$ 的綫性依賴性是有关的.

定理. 我們永远有 $|\rho(\mathbf{X}, \mathbf{Y})| \leqslant 1$，而且 $\rho(\mathbf{X}, \mathbf{Y}) = \pm 1$，仅当 $\mathbf{Y} = a\mathbf{X} + b$ 时才行，此处 a, b 都是常数.

証. 令 $\mathbf{X}^*$, $\mathbf{Y}^*$ 为正則化随机变量，則

$$\mathrm{Var}(\mathbf{X}^* \pm \mathbf{Y}^*) = \mathrm{Var}(\mathbf{X}^*) \pm 2\mathrm{Cov}(\mathbf{X}^*, \mathbf{Y}^*) + \mathrm{Var}(\mathbf{Y}^*) = $$
$$= 2(1 \pm \rho(\mathbf{X},\mathbf{Y})). \tag{8.2}$$

左边不可能为負，所以 $|\rho(\mathbf{X}, \mathbf{Y})| \leqslant 1$. 若 $\rho(\mathbf{X}, \mathbf{Y}) = 1$，則必有 $\mathrm{Var}(\mathbf{X}^* - \mathbf{Y}^*) = 0$，这意味差 $\mathbf{X}^* - \mathbf{Y}^*$ 只能取一个值. 即是 $\mathbf{X}^* - \mathbf{Y}^* =$ 常数. 故 $\mathbf{Y} = a\mathbf{X} +$ 常数，此处 $a = \sigma_y/\sigma_x$. 用同样推理可以証明 $\rho(\mathbf{X}, \mathbf{Y}) = -1$ 的情形.

9. 問 题

1. 七个球随机地分布在七个盒中。令 $\mathbf{X}_i$ 为恰巧包含 i 个球的盒数。应用第二章第 5 节中所列出的概率，写出 $(\mathbf{X}_2, \mathbf{X}_3)$ 的联合分布。

2. 擲两顆均匀的骰子。令 $\mathbf{X}$ 为第一顆骰子所出現的点数，而 $\mathbf{Y}$ 为两顆骰子中所出現的点数的最大者。(a) 写出 $(\mathbf{X}, \mathbf{Y})$ 的联合分布. (b) 求出均值，

方差，协方差.

3. 把一个錢币扔五次，令 $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ 分别为正面出現的次数；正面連貫的个数和最长的正面連貫的长度. 試把 32 个样本点和其对应的 $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ 的值在一起列一个表. 而且用簡单計算导出 $(\mathbf{X}, \mathbf{Y}), (\mathbf{X}, \mathbf{Z}), (\mathbf{Y}, \mathbf{Z})$ 之联合分布及 $\mathbf{X} + \mathbf{Y}, \mathbf{XY}$ 之分布. 求出这些随机变量之均值，方差，协方差.

4. 設随机变量 $\mathbf{X}_1$ 和 $\mathbf{X}_2$ 相互独立而且具有共同的几何分布 $\{q^k p\}$，其中 $k = 0, 1, \cdots$. 令 $\mathbf{Z}$ 为 $\mathbf{X}_1, \mathbf{X}_2$ 之中的較大者 [如果用符号来写的話，即是 $\mathbf{Z} = \max\{\mathbf{X}_1, \mathbf{X}_2\}$]. 推出 $\mathbf{Z}$ 和 $\mathbf{X}_1$ 的联合分布及 $\mathbf{Z}$ 的分布.

5. 令 $\mathbf{X}_1, \mathbf{X}_2$ 为独立随机变量序列，且分别具有普阿松分布 $\{p(k; \lambda_1)\}$，$\{p(k; \lambda_2)\}$. 証明 $\mathbf{X}_1 + \mathbf{X}_2$ 具有普阿松分布 $\{p(k; \lambda_1 + \lambda_2)\}$.

6（續上）. 在給定 $\mathbf{X}_1 + \mathbf{X}_2$ 的条件下 $\mathbf{X}_1$ 的条件概率分布是二項分布，即是

$$\mathbf{P}\{\mathbf{X}_1 = k | \mathbf{X}_1 + \mathbf{X}_2 = n\} = b\left(k; n, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right). \tag{9.1}$$

7. 令 $\mathbf{X}_1$ 和 $\mathbf{X}_2$ 相互独立且具有共同的几何分布 $\{q^k p\}$（如問題 4）. 不必計算，直接証明在給定 $\mathbf{X}_1 + \mathbf{X}_2$ 的条件下 $\mathbf{X}_1$ 的条件分布是均匀的，即是

$$\mathbf{P}\{\mathbf{X}_1 = k | \mathbf{X}_1 + \mathbf{X}_2 = n\} = \frac{1}{n+1}, \quad k = 0, 1, \cdots, n. \tag{9.2}$$

8. 令 $\mathbf{X}_1, \cdots, \mathbf{X}_n$ 为相互独立的随机变量，每一个都具有均匀分布 $\mathbf{P}\{\mathbf{X}_i = k\} = \frac{1}{N}$, $k = 1, 2, \cdots, N$. 令 $\mathbf{U}_n$ 和 $\mathbf{V}_n$ 分别 $\mathbf{X}_1, \cdots, \mathbf{X}_n$ 中之最小者和最大者. 求出 $\mathbf{U}_n$ 和 $\mathbf{V}_n$ 的分布. 它和估計問題(3.e)的联系是什么?

9. 在估計問題(3.e)中，求出最大和最小观察值的联合分布. 特別，$n=2$.（提示：首先計算 $\mathbf{P}\{\mathbf{X} \leqslant r, \mathbf{Y} \geqslant s\}$.）

10（續上）. 求出在条件 $\mathbf{X} = r$ 下最初两个观察值为 j 和 k 的条件概率.

11（續上）. 求出 $\mathbf{E}(\mathbf{X}^2)$，从而找出 $\mathrm{Var}(\mathbf{X})$ 当 $N \to \infty$ 的一个漸近表示式（n 固定）.

12. 抽样驗收. 假定一批产品中废品率为 q，而每一件被抽出来受检驗的概率为 p'. 因此，我們可以把全部产品分成四类，即是"接收了而且进行过检驗"，"接收了但是沒有进行过检驗"，等等，其对应的概率分别为 $pp', pq', p'q, qq'$，其中 $q = 1 - p$, $q' = 1 - p'$. 因此我們得到一串双重的伯努利試驗[参看第六章例(9.c)]. 令 $\mathbf{N}$ 为发現第一件废品以前通过检驗台的产品的件数（可以是眞正检驗过的，也可能是沒有检驗过的.），$\mathbf{K}$ 为其中沒有被

发現(沒有检驗)的废品的件数．求出 $\mathbf{N}$ 和 $\mathbf{K}$ 的联合分布及边緣分布．

13（續上）． 求 $\mathbf{E}\left(\frac{\mathbf{K}}{\mathbf{N}+1}\right)$ 和 $\mathrm{Cov}(\mathbf{K}, \mathbf{N})$．[在工业营业中，被发現的废品将要用合格产品去代替的，因此，$\mathbf{K}/(\mathbf{N}+1)$是其中废品的比例，它測量了这一批产品的質量，注意：$\mathbf{E}\left(\frac{\mathbf{K}}{\mathbf{N}+1}\right)$ 与 $\mathbf{E}(\mathbf{K})/\mathbf{E}(\mathbf{N}+1)$ 不一样．]

14．在一串伯努利試驗序列中，令 $\mathbf{X}$ 为从第一次試驗开始的連貫（成功連貫或失敗連貫）的长度．求出 $\mathbf{X}$ 的分布和 $\mathbf{E}(\mathbf{X})$，$\mathrm{Var}(\mathbf{X})$．

15（續上）． 令 $\mathbf{Y}$ 为第二个連貫的长度．求出 $\mathbf{Y}$ 的分布和 $\mathbf{E}(\mathbf{Y})$，$\mathrm{Var}(\mathbf{Y})$ 和 $\mathbf{X}, \mathbf{Y}$ 的联合分布．

16．如果 $\mathbf{X}, \mathbf{Y}$ 这两个随机变量中的每一个都只取两个值，而且 $\mathrm{Cov}(\mathbf{X}, \mathbf{Y})=0$，則 $\mathbf{X}$ 与 $\mathbf{Y}$ 是相互独立的．

17．**生日**． 对一个拥有 n 个人的人羣，求出一年中恰有 k 个人出生的那些日数的期望值．（一年取 365 天，而且一切排列都是等概的．）

18（續上）． 求出重复生日的日数的期望值．問 n 为多大时才能使这个期望值超过 1．

19．有一个人想开門．他共有 n 把鑰匙，可以这样設想，他用它們去試开門时所抽取的鑰匙是相互独立的而且是随机的． 求出試驗次数的均值与方差．(a)如果以前选出来的开不开門的鑰匙不除去；(b)与(a)相反．（假定只有一把鑰匙能打开这扇門．精确分布在第二章第 7 节中給出了，但現在这个問題并不要求精确分布．）

20．令 $(\mathbf{X}, \mathbf{Y})$ 为随机变量，其分布为 (1.9) 所給出的多項分布． 求出 $\mathbf{E}(\mathbf{X})$，$\mathrm{Var}(\mathbf{X})$，$\mathrm{Cov}(\mathbf{X}, \mathbf{Y})$：(a)直接計算；(b)把 $\mathbf{X}, \mathbf{Y}$ 都視为 n 个随机变量之和，并应用第 5 节的方法．

21．把一个骰子擲 n 次，求出么点出現的次数和六点出現的次数的协方差．

22．在第六章問題 24 的关于野兽陷井的問題中，証明在第 ν 次才陷入陷井的野兽的个数的期望值为 $nqp^{\nu-1}$．

23．如果 $\mathbf{X}$ 具有几何分布 $\mathbf{P}\{\mathbf{X}=k\}=q^k p$(其中 $k=0, 1, \cdots$)，証明 $\mathrm{Var}(\mathbf{X})=qp^{-2}$．当 r 为正整数时，推出負二項分布 $\{f(k; r, p)\}$ 具有方差 rqp^{-2}．用直接計算的方差証明上述事实对全部 $r>0$ 都对．

24．在(3.d)的等待时間問題中，証明

$$\mathrm{Var}(\mathbf{S}_r)=N\left\{\frac{1}{(N-1)^2}+\frac{2}{(N-2)^2}+\cdots+\frac{r-1}{(N-r+1)^2}\right\}.$$

提示：应用在問題 23 中所得到的关于几何分布的方差，附带地，当 $N\to\infty$ 时，我們发現 $N^{-2}\mathrm{Var}(\mathbf{S}_n)=\pi^2/6$。

25(續上)． 令 $\mathbf{Y}_r$ 为要求样本中包含 r 个指定的元素(代替課文中的任意 r 个不同的元素)所需之抽取次数． 求出 $\mathbf{E}(\mathbf{Y}_r)$ 和 $\mathrm{Var}(\mathbf{Y}_r)$．(注意：$\mathbf{Y}_r$ 的分布在第二章問題(11.12)中曾求出，不过現在这个問題并不要求用其分布．)

26[1]．N(相当大的数目)个人都去驗血．这可以用两种办法去进行．(i)每一个人都分别地去驗．这时，要求試驗 N 次． (ii)把从 k 个人中所抽出来的血混在一起进行分析，如果試驗是阴性的，那么对这 k 个人来說只作一次检驗就够了． 如果試驗是阳性的，那么必須对这 k 个人再逐个地分别試驗．这时对这 k 个人就須要作 $k+1$ 次試驗了．假定对所有的人来說，試驗是阳性反应的概率都是 p，而且这些人都是統計独立的．

(a) 从 k 个人中所抽出来的混合样本的試驗是阳性的概率是多少?

(b) 在方案(ii)下，所需要进行的試驗的次数 $\mathbf{X}$ 的期望值是多少?

(c) 为了使得在方案(ii)下所需要試驗的次数 $\mathbf{X}$ 的期望值最小，k 究竟应該怎样? 不必要对 k 进行数值的計算，因为这个問題导出了一个关于 k 的較繁的方程．

27．**样本結构**． 有一个由 r 类元素构成的总体，其中各类的元素的个数之比为 $p_1:p_2:\cdots:p_r$． 有放回地从这个总体中抽取一个大小为 n 的随机样本．求出在样本中不包含那些类的元素的类数的期望值．

28．令 $\mathbf{X}$ 为 r_1 个 α 和 r_2 个 β 的随机排列中 α-連貫的个数．在第二章問題(11.23)給出了 $\mathbf{X}$ 的分布．試求出 $\mathbf{E}(\mathbf{X})$和 $\mathrm{Var}(\mathbf{X})$．

29．在波利亚罐子模型中［第五章 (2.c)］，当第 n 次試驗的結果是黑时 $\mathbf{X}_n$ 为 1；結果是紅时 $\mathbf{X}$ 为 0，証明 $\rho(\mathbf{X}_n,\mathbf{X}_m)=c/(b+r+c)$，其中 $m\neq n$．

30(續上)． 令 $\mathbf{S}_n$ 为前 n 次抽取中黑球的总数(即是 $\mathbf{S}_n=\mathbf{X}_1+\mathbf{X}_2+\cdots+\mathbf{X}_n$)．求出 $\mathbf{E}(\mathbf{S}_n)$ 和 $\mathrm{Var}(\mathbf{S}_n)$．

31．**分层抽样**． 一个城市有 n 个区，其中住有 x_j 个居民的区共有 n_j 个 $(n_1+n_2+\cdots=n)$．令 $m=\sum n_jx_j/n$ 为每一区居民的平均数，且令 $a^2=\sum(n_jx_j^2/n)-m^2$．随机地无放回地选取 r 个区，然后把样本中每一区的居民的数目調查出来．令 $\mathbf{X}_1,\cdots,\mathbf{X}_r$ 分别为这 r 个区的居民数．試証

$$\mathbf{E}(\mathbf{X}_1+\cdots+\mathbf{X}_r)=mr,$$

1) 这个問題的来源是在第二次世界大战中所发展出来的一种新技术． 参看[69]．在軍队中实行方案 (ii) 能节省 80%．

$$\mathrm{Var}(\mathbf{X}_1+\cdots+\mathbf{X}_r)=\frac{a^2r(n-r)}{n-1}.$$

（注意：有放回的抽样的方差較大，其方差为 a^2r.）

32. 随机鏈的长度[1]. 在 xy 平面上，有一个由 n 个环所构成的鏈（每一环的长度为一个单位长）. 相邻二环之間的夹角各以 1/2 的概率取值 $\pm\alpha$，这里 α 为正常数. 又假定各夹角是相互独立的. 鏈的长度 L_n 是一个随机变量. 我們要証明

$$\mathbf{E}(\mathbf{L}_n^2)=n\frac{1+\cos\alpha}{1-\cos\alpha}-2\cos\alpha\frac{1-\cos^n\alpha}{(1-\cos\alpha)^2}. \tag{9.3}$$

不失普遍性，我們可以假定第一个环落在正 x 軸的方向，第 k 个环和正 x 軸的夹角为一随机变量 $\mathbf{S}_{k-1}$，这里 $\mathbf{S}_0=0$, $\mathbf{S}_k=\mathbf{S}_{k-1}+\alpha\mathbf{X}_k$. 而 $\mathbf{X}_k$ 是相互独立的随机变量，且各以 1/2 的概率取值 ±1. 第 k 个环在两个坐标軸上之投影分别为 $\cos\mathbf{S}_{k-1}$ 和 $\sin\mathbf{S}_{k-1}$. 所以对 $n\geqslant1$ 有

$$\mathbf{L}_n^2=\left(\sum_{k=0}^{n-1}\cos\mathbf{S}_k\right)^2+\left(\sum_{k=0}^{n-1}\sin\mathbf{S}_k\right)^2. \tag{9.4}$$

对 $m<n$ 用归納法依次地証明

$$\mathbf{E}(\cos\mathbf{S}_n)=\cos^n\alpha,\quad \mathbf{E}(\sin\mathbf{S}_n)=0, \tag{9.5}$$

$$\mathbf{E}((\cos\mathbf{S}_m)(\cos\mathbf{S}_n))=\cos^{n-m}\alpha\cdot\mathbf{E}(\cos^2\mathbf{S}_m), \tag{9.6}$$

$$\mathbf{E}((\sin\mathbf{S}_m)(\sin\mathbf{S}_n))=\cos^{n-m}\alpha\cdot\mathbf{E}(\sin^2\mathbf{S}_m), \tag{9.7}$$

$$\mathbf{E}(\mathbf{L}_n^2)-\mathbf{E}(\mathbf{L}_{n-1}^2)=1+2\cos\alpha\cdot\frac{1-\cos^{n-1}\alpha}{1-\cos\alpha} \tag{9.8}$$

$(L_0=0)$，从而最后得出(9.3).

33. 一串伯努利試驗序列一直継續到出現 r 次成功为止，其中 r 是一个固定的整数. 令 $\mathbf{X}$ 为需要的試驗次数. 求[2] $\mathbf{E}(r/\mathbf{X})$.（定义导出一个无穷級数，但它有有限的表达式.）

34. 在把 r 个球放入 n 个盒的随机排列中，恰巧发現 m 个盒是空的概率满足第二章的递推公式(11.8). 令 m_r 为空盒的个数的数学期望值. 从递推

1) 这是化学中长聚合物分子的长度問題的二維类似. 在我們的問題中随机变量 $\mathbf{L}_n$ 不能表为簡单的随机变量的和.

2) 这个例子說明随意停止与不随意停止的区別. 如果試驗的次数 n 固定，则成功次数 N 与試驗次数 n 之比是一个随机变量，其期望值为 p. 然而，在我們的例子中，当成功次数 r 固定，而所需之試驗次数 $\mathbf{X}$ 依賴于偶然性时，r 与 $\mathbf{X}$ 之比的期望值就不一定是 p 了. 例如，$p=\frac{1}{2}$ 时，对 $r=2$ 我們有 $\mathbf{E}(2/\mathbf{X})=0.614$，而不是 0.5，对 $r=3$，我們有 $\mathbf{E}(3/\mathbf{X})=0.579$.

公式証明

$$m_{r+1}=1+(1-n^{-1})m_r,$$

并且推出

$$m_r=n\left\{1-\left(1-\frac{1}{n}\right)^r\right\}.$$

35.令 $\mathbf{S}_n$ 为 n 次伯努利試驗的成功次数。証明

$$\mathbf{E}(|\mathbf{S}_n-np|)=2\nu qb(\nu;n,p),$$

其中 ν 为滿足 $np<\nu\leqslant np+1$ 的整数。

36.令$\{\mathbf{X}_k\}$为相互独立且具有共同分布的随机变量序列。假定 $\mathbf{X}_k$ 只取正值而且 $\mathbf{E}(\mathbf{X}_k)=a$ 和 $\mathbf{E}(\mathbf{X}_k^{-1})=b$ 存在。令 $\mathbf{S}_n=\mathbf{X}_1+\cdots+\mathbf{X}_n$.証明 $\mathbf{E}(\mathbf{S}_n^{-1})$ 是有限的，而且 $\mathbf{E}(\mathbf{X}_k/\mathbf{S}_n)=\frac{1}{n}$ 对 $k=1,2,\cdots,n$.

37(續上)[1]. 証明

$$\mathbf{E}\left(\frac{\mathbf{S}_m}{\mathbf{S}_n}\right)=\frac{m}{n},\ \text{如果 } m\leqslant n,$$

$$\mathbf{E}\left(\frac{\mathbf{S}_m}{\mathbf{S}_n}\right)=1+(m-n)a\mathbf{E}(\mathbf{S}_n^{-1}),\ \text{如果 } m\geqslant n.$$

38.令 $\mathbf{X}_1,\cdots,\mathbf{X}_n$ 为相互独立且具有共同分布的随机变量，令其均值为 m，方差为 σ^2. 令 $\overline{\mathbf{X}}=(\mathbf{X}_1+\cdots+\mathbf{X}_n)/n$. 証明[2]:

$$\frac{1}{n-1}\mathbf{E}\left(\sum_{k=1}^{n}(\mathbf{X}_k-\overline{\mathbf{X}})^2\right)=\sigma^2.$$

39.令 $\mathbf{X}_1,\cdots,\mathbf{X}_n$ 是相互独立的随机变量。令 $\mathbf{U}$ 为 $\mathbf{X}_1,\cdots,\mathbf{X}_k$ 的函数，$\mathbf{V}$ 为 $\mathbf{X}_{k+1},\cdots,\mathbf{X}_n$ 的函数($k<n$). 証明 $\mathbf{U},\mathbf{V}$ 是相互独立的随机变量.

40.切比謝夫不等式的推广. 令 $\phi(x)$ 在 $x>0$ 的地方是一个单調上升的正函数而且假定 $\mathbf{E}(\phi(|\mathbf{X}|))=M$ 存在。証明

$$\mathbf{P}\{|\mathbf{X}|\geqslant t\}\leqslant\frac{M}{\phi(t)}.$$

41.希瓦茲 (Schwarz) 不等式. 对于任何两个具有有限方差的随机变量 $\mathbf{X}$ 和 $\mathbf{Y}$ 来說，都有 $\mathbf{E}^2(\mathbf{XY})\leqslant\mathbf{E}(\mathbf{X}^2)\mathbf{E}(\mathbf{Y}^2)$. 由二次多項式 $\mathbf{E}((t\mathbf{X}+\mathbf{Y})^2)$ 非負这一事实証明上述不等式.

---

1) 习題 37 可由习題 36 推出的意見是由鍾开萊提出的.
2) 这可解释为: $\Sigma(\mathbf{X}_k-\overline{\mathbf{X}})^2/(n-1)$ 是 σ^2 之无偏估计量.

問 題 解 答

第 一 章

1.(a) 3/5; (b) 3/5; (c) 3/10.

2.事件 S_1, S_2, $S_1 \cup S_2$ 以及 S_1S_2 分别包含了 12, 12, 18 和 6 个点.

4.这个空間包含了两点 HH 和 TT,其对应的概率各为 1/4; 包含了两点 HTT 和 THH, 其对应的概率各为 1/8; 一般地包含各具有概率 2^{-n} 的两个点 ($n \geqslant 2$). 这些概率之和为 1, 因此不需要考虑无穷的抛擲的序列的概率. 我們所要求的两个概率分别为 15/16 和 2/3.

9.$\mathbf{P}\{AB\} = 1/6$, $\mathbf{P}\{A \cup B\} = 23/36$, $\mathbf{P}\{AB'\} = 1/3$.

12.在事件(a), (b)和(g)中, $x = 0$.

在事件(e), (f)中, $x = 1$.

在事件(d)中, $x = 2$.

在事件(c)中, $x = 4$.

15.(a) A; (b) AB; (c) $B \cup (AC)$.

16.(c), (d), (e), (f), (h), (i), (k), (l) 是对的. 除非 $C \subset B$, (a) 是无意义的. 甚至在这一情形下, 一般也是不对的, 但是在特殊情形 $C \subset B$, $AC = 0$ 下,它是对的. 如果 $C \supset AB$, 则(b)是对的. (g)应該是$(A \cup B) - A = A'B$. 最后,由于(k)是对的故(j)不对.

17.(a) $AB'C'$; (b) ABC'; (c) ABC; (d) $A \cup B \cup C$; (e) $AB \cup AC \cup BC$; (f) $AB'C' \cup A'BC' \cup A'B'C$; (g) $ABC' \cup AB'C \cup A'BC = (AB \cup AC \cup BC) - ABC$; (h) $A'B'C'$; (i) $(ABC)'$.

18.$A \cup B \cup C = A \cup (B - AB) \cup \{C - C(A \cup B)\} = A \cup BA' \cup CA'B'$.

第 二 章

1.(a) 26^3; (b) $26^2 + 26^3 = 18,252$; (c) $26^2 + 26^3 + 26^4$.

2.$64 \cdot 14 = 896$. 对具有 n^2 个区域的棋盘来說,公式为 $n^2(2n - 2)$.

3.$2(2^{10} - 1) = 2046$.

4. $\binom{n}{2}+n=\frac{n(n+1)}{2}$.

5. (a) $\frac{1}{n}$; (b) $\frac{1}{n(n-1)}$.

6. (a) $p_1=0.01$, $p_2=0.27$, $p_3=0.72$; (b) $p_1=0.001$, $p_2=0.063$, $p_3=0.432$, $p_4=0.504$.

7. $p_r=(10)_r 10^{-r}$. 例如 $p_3=0.72$, $p_{10}=0.00036288$. 史特令漸近公式給出 $p_{10}=0.0003598\cdots$.

8. (a) $(9/10)^k$; (b) $(9/10)^k$; (c) $(8/10)^k$; (d) $2(9/10)^k-(8/10)^k$; (e) AB 和 $A\cup B$.

9. $\binom{n}{2}n!n^{-n}$.

10. $9\div\binom{12}{8}=\frac{1}{55}$.

11. 恰为 r 次試驗的概率为 $(n-1)_{r-1}\div(n)_r=n^{-1}$.

12. (a) $1/1\cdot3\cdot5\cdots(2n-1)=2^n n!/(2n)!$; (b) $(n!)/1\cdot3\cdots(2n-1)=2^n/\binom{2n}{n}$.

13. 假定它具有随机性，则 12 次接見訪問全部都在星期二或者在星期四的概率为 $(2/7)^{12}=0.0000003\cdots$. 只有 $\binom{7}{2}=21$"对天"，所以即使对任意两天来説，这个概率仍然很小. 因此，有理由說接見訪問的日子是有一定的規定的.

14. 假定它具有随机性，此事件的概率是 $\left(\frac{6}{7}\right)^{12}$，近似于 $\frac{1}{6}$. 沒有可靠的結論是可能的.

15. $(90)_{10}\div(100)_{10}=0.330476\cdots$.

16. $(25!)(5!)^{-5}5^{-25}=0.00209\cdots$.

17. $\frac{2(n-2)_r(n-r-1)!}{n!}=\frac{2(n-r-1)}{n(n-1)}$.

18. (a) $\frac{1}{216}$; (b) $\frac{83}{3888}$.

19. 概率为 $1-\left(\frac{5}{6}\right)^4=0.517747\cdots$ 和 $1-\left(\frac{35}{36}\right)^{24}=0.491404\cdots$.

20. (a) $(n-N)_r\div(n)_r$; (b) $(1-N/n)^r$. 对于 $r=N=3$ 时，概率为 (a) $0.911812\cdots$; (b) $0.912673\cdots$. 对于 $r=N=10$，它們是 (a) 0.330476;

(b)0.348678⋯.

21. (a) $(1-N/n)^{r-1}$; (b) $(n)_{Nr}\div((n)_N)^r$.

22. $(1-2/n)^{2r-2}$; 中位数近似地等于 $2^{r+1}=0.7n$.

23. 假定它具有随机性，三个或者四个都是(a)某一个女孩打破的；(b)最年輕的女孩打破的概率分别为$\frac{13}{64}\approx 0.2$ 和 $\frac{13}{256}\approx 0.05$.

24. (a) $12!/12^{12}=0.000054$; (b) $\binom{12}{2}(2^6-2)12^{-6}=0.00137\cdots$.

25. $\frac{30!}{2^6 6^6}\binom{12}{6}12^{-30}\approx 0.00035\cdots$.

26. (a) $\binom{n}{2r}2^{2r}\div\binom{2n}{2r}$; (b) $n\binom{n-1}{2r-2}2^{2r-2}\div\binom{2n}{2r}$;

(c) $\binom{n}{2}\binom{n-2}{2r-4}2^{2r-4}\div\binom{2n}{2r}$.

27. $\binom{N-3}{r-1}\div\binom{N-1}{r-1}$.

28. $p=\binom{2N}{N}^2\div\binom{4N}{2N}\approx\{2/N\pi\}^{1/2}$.

29. $p=\dfrac{\binom{4}{k}\binom{48}{13-k}\binom{39}{13}\binom{26}{13}}{\binom{52}{13}\binom{39}{13}\binom{26}{13}}=\dfrac{\binom{4}{k}\binom{48}{13-k}}{\binom{52}{13}}$.

30. 参看問題 29. 其概率为

$$\binom{13}{m}\binom{39}{13-m}\binom{13-m}{n}\binom{26+m}{13-n}\div\binom{52}{13}\binom{39}{13}.$$

31. $\binom{4}{k}\binom{48}{26-k}\div\binom{52}{26}$.

32. $\dfrac{\binom{13}{a}\binom{39}{13-a}\binom{13-a}{b}\binom{26+a}{13-b}\binom{13-a-b}{c}\binom{13+a+b}{13-c}}{\binom{52}{13}\binom{39}{13}\binom{26}{13}}$.

33. (a) $24p(5,4,3,1)$; (b) $4p(4,4,4,1)$; (c) $12p(4,4,3,2)$.

34. $\dfrac{\binom{13}{a}\binom{13}{b}\binom{13}{c}\binom{13}{d}}{\binom{52}{13}}$. [关于一付牌含有某种花色的牌 a 张，另一种的 b 张，等等，这一事件的概率，参看問题 33.]

35. $p_0(r)=(52-r)_4\div(52)_4$; $p_1(r)=4r(52-r)_3\div(52)_4$; $p_2(r)=6r(r-1)(52-r)_2\div(52)_4$; $p_3(r)=4r(r-1)(r-2)(52-r)\div(52)_r$;

$p_4(r)=(r)_4\div(52)_4$.

36.第 1，…，第 4 个爱司的等待时間超过 r 的概率为 $w_1(r)=p_0(r)$；$w_2(r)=p_0(r)+p_1(r)$；$w_3(r)=p_0(r)+p_1(r)+p_2(r)$；$w_4(r)=1-p_4(r)$。其次，$f_i(r)=w_i(r)-w_i(r+1)$．中位数为 8，20，32，44。

37. (a) $\binom{4}{k}\binom{4-k}{k}\binom{48}{r-k}\binom{48-r+k}{r-k}\div\binom{52}{r}\binom{52-r}{r}$，其中 $k\leqslant 2$；

(b) $\left\{\binom{4}{k}\binom{48}{r-k}\div\binom{52}{r}\right\}^2$，其中 $k\leqslant 4$。

39. $\binom{r_1+n-1}{r_1}\binom{r_2+n-1}{r_2}$.

40. $\binom{r_1+5}{5}(r_2+1)$.

41. $\dfrac{(r_1+r_2+r_3)!}{r_1!r_2!r_3!}$.

42. $(49)_4\div(52)_4$.

43. $\mathbf{P}\{(7)\}$

$=10\cdot 10^{-7} \qquad =0.000\ 001.$

$\mathbf{P}\{(6,1)\}$

$=10\cdot 9\cdot\binom{7}{6}10^{-7} \qquad =\dfrac{10!}{8!1!1!}\cdot\dfrac{7!}{1!6!}\cdot 10^{-7} \qquad =.000\ 063.$

$\mathbf{P}\{(5,2)\}$

$=10\cdot 9\cdot\binom{7}{5}10^{-7} \qquad =\dfrac{10!}{8!1!1!}\cdot\dfrac{7!}{2!5!}\cdot 10^{-7} \qquad =.000\ 189.$

$\mathbf{P}\{(5,1,1)\}$

$=10\cdot\binom{9}{2}\binom{7}{5}2\cdot 10^{-7} \qquad =\dfrac{10!}{7!2!1!}\cdot\dfrac{7!}{1!1!5!}\cdot 10^{-7} \qquad =.001\ 512.$

$\mathbf{P}\{(4,3)\}$

$=10\cdot 9\cdot\binom{7}{4}\cdot 10^{-7} \qquad =\dfrac{10!}{8!1!1!}\cdot\dfrac{7!}{3!4!}\cdot 10^{-7} \qquad =.000\ 315.$

$\mathbf{P}\{(4,2,1)\}$

$=10\cdot 9\cdot 8\cdot\binom{7}{4}\binom{3}{2}\cdot 10^{-7} \qquad =\dfrac{10!}{7!1!1!1!}\cdot\dfrac{7!}{1!2!4!}\cdot 10^{-7} \qquad =.007\ 560.$

$\mathbf{P}\{(4,1,1,1)\}$

$=10\cdot\binom{9}{3}\binom{7}{4}3\cdot 2\cdot 10^{-7} \qquad =\dfrac{10!}{6!3!1!}\cdot\dfrac{7!}{1!1!1!4!}\cdot 10^{-7} \qquad =.017\ 640.$

$\mathbf{P}\{(3,3,1)\}$

$$=\binom{10}{2}8\cdot\binom{7}{3}\binom{4}{3}\ 10^{-7} \qquad =\frac{10!}{7!2!1!}\cdot\frac{7!}{1!3!3!}\cdot 10^{-7} \qquad = .005\ 040.$$

$\mathbf{P}\{(3,2,2)\}$

$$=10\binom{9}{2}\binom{7}{3}\binom{4}{2}\cdot 10^{-7} \qquad =\frac{10!}{7!2!1!}\cdot\frac{7!}{2!2!3!}\cdot 10^{-7} \qquad = .007\ 560.$$

$\mathbf{P}\{(3,2,1,1)\}$

$$=10\cdot 9\binom{8}{2}\binom{7}{3}\binom{4}{2}2\cdot 10^{-7} \qquad =\frac{10!}{6!2!1!1!}\cdot\frac{7!}{1!1!2!3!}\cdot 10^{-7} = .105\ 840.$$

$\mathbf{P}\{(3,1,1,1,1)\}$

$$=10\binom{9}{4}\binom{7}{3}4\cdot 3\cdot 2\cdot 10^{-7} \qquad =\frac{10!}{5!4!1!}\cdot\frac{7!}{1!1!1!1!3!}\cdot 10^{-7} \qquad = .105\ 840.$$

$\mathbf{P}\{(2,2,2,1)\}$

$$=\binom{10}{3}7\binom{7}{2}\binom{5}{2}\binom{3}{2}\cdot 10^{-7} \qquad =\frac{10!}{6!3!1!}\cdot\frac{7!}{1!2!2!2!}\cdot 10^{-7} \qquad = .052\ 920.$$

$\mathbf{P}\{(2,2,1,1,1)\}$

$$=\binom{10}{2}\binom{8}{3}\binom{7}{2}\binom{5}{2}3\cdot 2\cdot 10^{-7} \qquad =\frac{10!}{5!3!2!}\cdot\frac{7!}{1!1!1!2!2!}\cdot 10^{-7} = .317\ 520.$$

$\mathbf{P}\{(2,1,1,1,1,1)\}$

$$=10\binom{9}{5}\binom{7}{2}5\cdot 4\cdot 3\cdot 2\cdot 10^{-7} \qquad =\frac{10!}{4!5!1!}\cdot\frac{7!}{1!1!1!1!1!2!}\cdot 10^{-7} = .317\ 520.$$

$\mathbf{P}\{(1,1,1,1,1,1,1)\}$

$$=\binom{10}{7}7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 10^{-7} \qquad =\frac{10!}{3!7!}\cdot 7!\cdot 10^{-7} \qquad = .060\ 480.$$

44.令 S, D, T, Q 分別表示一重,二重,三重,四重,則我們有

$$\mathbf{P}\{22S\} \qquad = \frac{365!}{22!343!}\cdot 365^{-22} \qquad = 0.524\ 30.$$

$$\mathbf{P}\{20S+1D\} \qquad = \frac{365!}{20!1!344!}\cdot\frac{22!}{20!2!}\cdot 365^{-22} \qquad = .352\ 08.$$

$$\mathbf{P}\{18S+2D\} \qquad = \frac{365!}{18!2!345!}\cdot\frac{22!}{18!2!2!}\cdot 365^{-22} \qquad = .096\ 95.$$

$$\mathbf{P}\{16S+3D\} \qquad = \frac{365!}{16!3!346!}\cdot\frac{22!}{16!2!2!2!}\cdot 365^{-22} \qquad = .014\ 29.$$

$$\mathbf{P}\{19S+1T\} \qquad = \frac{365!}{19!1!345!}\cdot\frac{22!}{19!3!}\cdot 365^{-22} \qquad = .006\ 80.$$

$$\mathbf{P}\{17S+1D+1T\} = \frac{365!}{17!1!1!346!}\cdot\frac{22!}{19!3!}\cdot 365^{-22} = .003\ 36.$$

$$\mathbf{P}\{14S+4D\} = \frac{365!}{14!4!347!}\cdot\frac{22!}{14!2!2!2!2!}\cdot 365^{-22} = .001\ 24.$$

$$\mathbf{P}\{15S+2D+1T\} = \frac{365!}{15!2!1!347!}\cdot\frac{22!}{15!2!2!3!}\cdot 365^{-22} = .000\ 66.$$

$$\mathbf{P}\{18S+1Q\} = \frac{365!}{18!1!346!}\cdot\frac{22!}{18!4!}\cdot 365^{-22} = .000\ 09.$$

45.令 $q=\binom{52}{5}=2,598,960$。概率为(a)$4/q$; (b)$13\cdot 12\cdot 4\cdot q^{-1}=\frac{1}{4165}$; (c) $13\cdot 12\cdot 4\cdot 6\cdot q^{-1}=\frac{6}{4165}$; (d)$9\cdot 4^5\cdot q^{-1}=\frac{768}{216580}$; (e)$13\cdot\binom{12}{2}4\cdot 4^2\cdot q^{-1}=\frac{88}{4165}$; (f)$\binom{13}{2}\cdot 11\cdot 6\cdot 6\cdot 4\cdot q^{-1}=\frac{198}{4165}$; (g)$13\cdot\binom{12}{3}\cdot 6\cdot 4^3\cdot q^{-1}=\frac{1760}{4165}$.

第　四　章

1.99/323.

2.0.21….

3.1/4.

4.$7/2^6$.

5.1/81 和 $31/6^6$.

6.若 A_k 为事件(k, k)不出現,則由(1.5)有

$$1-p_r=6\left(\frac{35}{36}\right)^r-\binom{6}{2}\left(\frac{34}{36}\right)^r+\binom{6}{3}\left(\frac{33}{36}\right)^r-\binom{6}{4}\left(\frac{32}{36}\right)^r+6\left(\frac{31}{36}\right)^r-\left(\frac{30}{36}\right)^r.$$

7.令 $p^{-1}=\binom{52}{13}$,則 $S_1=13\binom{48}{9}p$; $S_2=\binom{13}{2}\binom{44}{5}p$; $S_3=40\cdot\binom{13}{3}\cdot p$. 近似值为 $P_{[0]}=0.9658$; $P_{[1]}=0.0341$; $P_{[2]}=0.0001$.

8. $u_r=\sum_{k=0}^{N}(-1)^k\binom{N}{k}\left(1-\frac{k}{n}\right)^r$.

9. $p_r=\sum_{k=0}^{N}(-1)^k\binom{N}{k}\frac{(n-k)_r}{(n)_r}$. 关于两个公式的一致性的証明,請参看第二章(12.18).

10.一般項为 $a_{1k_1}a_{2k_2}\cdots a_{Nk_N}$,其中$(k_1, k_2, \cdots, k_N)$为$(1, 2, \cdots, N)$的一个排列。对于对角元素来說 $k_\nu=\nu$.

12. $u_r = \sum_{k=0}^{n}(-1)^k\binom{n}{k}\frac{(ns-ks)_r}{(ns)_r}$.

14. 注意：由定义知 $r<n$ 时 $u_r=0$，且 $u_n = n!s^n/(ns)_n$。

15. $u_r - u_{r-1} = \sum_{k=1}^{n}(-1)^{k-1}\binom{n-1}{k-1}\frac{(ns-ks)_{r-1}}{(ns-1)_{r-1}} \to$

$$\to \sum_{k=0}^{n-1}(-1)^k\binom{n-1}{k}\left(1-\frac{k+1}{n}\right)^{r-1}.$$

16. $\binom{N}{2}^{-r}\binom{N}{m}\sum_{k=2}^{m}(-1)^{m-k}\binom{m}{k}\binom{k}{2}^r$.

17. 应用 $\binom{52}{5}S_k = \binom{4}{k}\binom{52-13k}{5}$. 则近似地有 $P_{[0]} = 0.264$，$P_{[1]} = 0.588$，$P_{[2]} = 0.146$，$P_{[3]} = 0.002$。

18. 应用 $\binom{52}{13}$ $S_k = \binom{4}{k}\binom{52-2k}{13-2k}$. 则近似地有 $P_{[0]} = 0.780217$，$P_{[1]} = 0.204606$，$P_{[2]} = 0.014845$，$P_{[3]} = 0.000330$，$P_{[4]} = 0.000002$。

19. $m!N!u_m = \sum_{k=0}^{N-m}(-1)^k(N-m-k)!/k!$.

20. 参看下面的公式(对 $r=2$ 的情形)。

21. $(rN)!x = \binom{N}{2}r^2(rN-2)! - \binom{N}{3}r^3(rN-3)! + -\cdots +$

$$+ (-1)^N r^N(rN-N)!.$$

24. $P_{[m]} = \frac{\binom{n}{m}}{\binom{n+r-1}{r}}\sum_{k=0}^{n-m}(-1)^k\binom{n-m}{k}\binom{n-m+r-1-k}{r}$.

25. 利用第二章(12.16)和(12.4)。

26. 令 $U_N = A_1 \cup \cdots \cup A_N$ 并注意 $U_{N+1} = U_N \cup A_{N+1}$ 和 $U_N A_{N+1} = (A_1 A_{N+1}) \cup \cdots \cup (A_N A_{N+1})$。

第　五　章

1. $1 - \frac{(5)_3}{(6)_3} = \frac{1}{2}$.

2. $p = 1 - \frac{10\cdot 5^9}{6^{10}-5^{10}} = 0.61\cdots$.

3. (a) $\binom{35}{13} \div \binom{39}{13} = 0.182\cdots$，恰有一个爱司的概率为 $4\cdot\binom{35}{12}\div\binom{39}{13}=$

$0.411\cdots$; (b)近似地为 $1-0.182-0.411\approx 0.407$。

4. (a) $2\cdot\frac{\binom{23}{10}}{\binom{26}{13}}=\frac{11}{50}$; (b)$2\cdot\frac{\binom{23}{12}}{\binom{26}{13}}=\frac{13}{50}$.

6. $\frac{125}{345}$; $\frac{140}{345}$; $\frac{80}{345}$.

7. $\frac{20}{21}$.

9. $\left(\frac{5}{6}\right)^2$.

10. $1-\left(\frac{5}{6}\right)^2$.

12. $\frac{p}{2-p}$.

13. (b)$\frac{3}{5}$; (c)$2^n\cdot(1+2^n)^{-1}$.

14. (d) 令 $a_n=x_n-\frac{4}{7}$, $b_n=y_n-\frac{1}{7}$, $c_n=z_n-\frac{2}{7}$, 則 $|a_n|+|b_n|+|c_n|=\frac{1}{2}\{|a_{n+1}|+|b_{n+1}|+|c_{n+1}|\}$. 因此, $|a_n|+|b_n|+|c_n|$ 几何級数地上升.

15. $p=(1-p_1)(1-p_2)\cdots(1-p_n)$.

16. 应用不等式 $1-x<e^{-x}(0<x<1)$, 或者 $\log(1-x)$ 的泰劳級数; 参看第二章(8.12).

18. $\frac{b+c}{b+c+r}$.

19. 如果不管 b, r 和 c 这个事实对第 n 次抽取是对的話, 则在第 $n+1$ 次試驗是紅的概率为

$$\frac{b}{b+r}\cdot\frac{b+c}{b+r+c}+\frac{r}{b+r}\cdot\frac{b}{b+r+c}=\frac{b}{b+r}.$$

20. 前面一个問題說明,这个論断对 $m=1$ 和一切 n 都是正确的. 由归納法,考虑第一次試驗的两种可能性.

23. 利用第二章(12.9).

26. 从(5.2) $2v=2p(1-p)\leq\frac{1}{2}$.

28. (a)u^2; (b)$u^2+uv+v^2/4$; (c)$u^2+(25uv+9v^2+vw+2uw)/16$.

33. $p_{11}=p_{32}=2p_{21}=p$, $p_{12}=p_{33}=2p_{23}=q$, $p_{13}=p_{31}=0$, $p_{22}=\frac{1}{2}$.

第　六　章

1. 5/16.

2. 其概率为0.02804….

3. $(0.9)^x\leqslant 0.1, x\geqslant 22$.

4. $q^x\leqslant\frac{1}{2}$ 和 $(1-4q)^x\leqslant\frac{1}{2}$,其中 $p=\binom{48}{9}\div\binom{52}{13}$.因此分别为$x\geqslant 263$ 和 $x\geqslant 66$.

5. $1-(0.8)^{10}-2(0.8)^9=0.6242$.

6. $\{1-(0.8)^{10}-2(0.8)^9\}/\{1-(0.8)^{10}\}=0.6993$.

7. $\binom{26}{2}\binom{26}{11}\div\binom{52}{13}=0.003954\cdots$ 和 $\binom{13}{2}\frac{1}{2^{13}}=0.00952\cdots$.

8. $\binom{12}{2}\{6^{-6}-2\cdot 12^{-6}\}$.

9. 正确的值为 0.6651…, 0.40187… 和 0.2009…; 普阿松逼近为 $1-e^{-1}=0.6321\cdots$, 0.3679… 和 0.1839….

10. $e^{-2}\sum_{4}^{\infty}2^k/k!=0.143\cdots$.

11. $e^{-1}\sum_{3}^{\infty}1/k!=0.080\cdots$.

12. $e^{-x/100}\leqslant 0.05$ 或者 $x\geqslant 300$.

13. $e^{-1}=0.3679\cdots$, $1-2\cdot e^{-1}=0.264\cdots$.

14. $e^{-x}\leqslant 0.01$, $x\geqslant 5$.

15. $1/p=649,740$.

16. $1-p^n$, 其中 $p=p(0;\lambda)+\cdots+p(k;\lambda)$.

18. 当 $k=0$ 为 q^3; 当 $k=1,2,3$ 时为 pq^3; 当 $k=4$ 时为 pq^3-pq^6.

19. $\sum_{k=0}^{n}\binom{n}{k}^2 2^{-2n}=\binom{2n}{n}2^{-2n}\approx\left(\frac{1}{\pi n}\right)^{1/2}$ 对充分大的 n.

20. $\sum_{k=a}^{a+b-1}\binom{a+b-1}{k}p^kq^{a+b-1-k}$. 这也可以写成另一形式:

$p^a\sum_{k=0}^{b-1}\binom{a+k-1}{k}q^k$，其中第 k 项是第 a 次成功发生在第 $k\leqslant b-1$ 次失敗以后的概率。

21. $x_r=\binom{2N-1-r}{N-1}\cdot 2^{-2N+r+1}$.

22. (a) $x=\sum_{r=1}^{N}x_r 2^{-r-1}=2^{-2N}\sum_{r=1}^{N}\binom{2N-1-r}{N-1}$；(b)应用第二章(12.6)。

23. $k_i\approx np_i$；$k_{12}\approx np_1p_2$ 这时 $n\approx k_1k_2/k_{12}$。

24. $\binom{n}{n_1}\binom{n-s_1}{n_2}\cdots\binom{n-s_{r-1}}{n_r}\cdot q^{s_r}p^{(rn-s_1-\cdots-s_r)}$，其中 $s_i=n_1+\cdots+n_i$。

25. $p=p_1q_2(p_1q_2+p_2q_1)^{-1}$.

31. 由对数的泰劳展开

$$b(0;n,p)=q^n=(1-\lambda/n)^n<e^{-\lambda}=p(0,\lambda).$$

对每一个分布来說其各項之和为 1，因此，不可能有这样一个分布，它的各項都大于另一个分布的对应項。

32. 普阿松分布中只有有限多項大于 ε，其余各項支配二項分布的对应的項。

第　七　章

1. 如第 1 节一样进行。

2. 应用(1.7)。

3. $\Phi\left(-\frac{32}{30}\right)=0.143\cdots$.

4. 0.99.

5. 500.

6. 66,400.

7. 更确切地。 第六章的不等式足以說明：超过 8 倍标准差是很不可能的。

8. $(2\pi n)^{-1}\{p_1p_2(1-p_1-p_2)\}^{-1/2}$.

第　八　章

1. $\beta=21$.

2. $x=pu+qv+rw$，其中 u,v,w 为下列方程組的解：

$$u = p^{\alpha-1} + (qv + rw)\frac{1 - p^{\alpha-1}}{1 - p}, \quad v = (pu + rw)\frac{1 - q^{\beta-1}}{1 - q},$$

$$w = pu + qv + rw = x.$$

3. $u = p^{\alpha-1} + (qv + rw)\dfrac{1 - p^{\alpha-1}}{1 - p}$,

$$v = (pu + rw)\frac{1 - q^{\beta-1}}{1 - q}, \quad w = (pu + qv)\frac{1 - r^{\gamma-1}}{1 - r}.$$

4. 注意：$\mathbf{P}\{A_n\} < (2p)^n$，但是

$$\mathbf{P}\{A_n\} > 1 - (1 - p^n)^{2^n/2n} > 1 - e^{-(2p)^n/2n}.$$

如果 $p = \dfrac{1}{2}$，则最后一个量$\sim 1/2n$；如果 $p > 1$，则 $\mathbf{P}\{A_n\}$ 不会趋于 0.

第　九　章

1. 可能的組合为$(0,0)$，$(0,1)$，$(0,2)$，$(1,0)$，$(1,1)$，$(2,0)$，$(2,1)$，$(3,0)$. 它們的概率分别为 0.047539，0.108883，0.017850，0.156364，0.214197，0.321295，0.026775，0.107098.

2. (a)联合分布可以写成 6 行 6 列的矩陣. 其主对角綫上的元素为 q，$2q$，…，$6q$，其中 $q = \dfrac{1}{36}$. 主对角綫一边的元素全为 0，另一边的元素全为 q. (b)$\mathbf{E}(\mathbf{X}) = 7/2$，$\text{Var}(\mathbf{X}) = \dfrac{35}{12}$，$\mathbf{E}(\mathbf{Y}) = \dfrac{161}{36}$，$\text{Var}(\mathbf{Y}) = \dfrac{2555}{1296}$，$\text{Cov}(\mathbf{X}, \mathbf{Y}) = \dfrac{105}{72}$.

3. 在 $\mathbf{X}, \mathbf{Y}$ 的联合分布中，其各行为以 32^{-1} 乘$(1, 0, 0, 0, 0, 0)$，$(0, 5, 4, 3, 2, 1)$，$(0, 0, 6, 6, 3, 0)$，$(0, 0, 0, 1, 0, 0)$；

对 $\mathbf{X}, \mathbf{Z}$ 的联合分布来說，其各行为 32^{-1} 乘$(1, 0, 0, 0, 0, 0)$，$(0, 5, 6, 1, 0, 0)$，$(0, 0, 4, 6, 1, 0)$，$(0, 0, 0, 3, 2, 0)$，$(0, 0, 0, 0, 2, 0)$，$(0, 0, 0, 0, 0, 1)$；

对 $\mathbf{Y}, \mathbf{X}$ 的联合分布来說，其各行为 32^{-1} 乘$(1, 0, 0, 0)$，$(0, 5, 6, 1)$，$(0, 4, 7, 0)$，$(0, 3, 2, 0)$，$(0, 2, 0, 0)$，$(0, 1, 0, 0)$.

$\mathbf{X} + \mathbf{Y}$ 的分布为$(1, 0, 5, 4, 9, 8, 5)$ 除以 32，而 $\mathbf{X} + \mathbf{Y}$ 的值从 0 到 6；$\mathbf{XY}$ 的分布为 $(1, 5, 4, 3, 8, 1, 6, 0, 3, 1)$ 除以 32，$\mathbf{XY}$ 的值由 0 到 9. $\mathbf{E}(\mathbf{X}) = 5/2$，$\mathbf{E}(\mathbf{Y}) = 3/2$，$\mathbf{E}(\mathbf{Z}) = 31/16$，$\text{Var}(\mathbf{X}) = 5/4$，$\text{Var}(\mathbf{Y}) = 3/8$，$\text{Var}(\mathbf{Z}) = 303/256$.

4. $\mathbf{P}\{\mathbf{Z} = i, \mathbf{X} = j\} = q^{i+j}p^2, i > j$；$\mathbf{P}\{\mathbf{Z} = i, \mathbf{X} = j\} = (1 - q^{i+1})q^i p$，$i = j$；

其它的值是不可能的。 $\mathbf{P}\{\mathbf{Z}=i\}=2q^ip-q^{2i}p-q^{2i+1}p$.

8. $\mathbf{V}_n$ 的分布由(3.5)所給出由于对称性，$\mathbf{U}_n$ 的分布同样可求出。

9. $\mathbf{P}\{\mathbf{X}\leqslant r,\mathbf{Y}\geqslant s\}=\left(\frac{r-s+1}{N}\right)^n$ 对 $r\geqslant s$;

$$\mathrm{P}\{\mathbf{X}=r,\mathbf{Y}=s\}=\begin{cases}N^{-n}\{(r-s+1)^n-2(r-s)^n+(r-s-1)^n\}, & r>s\\ N^{-n}, & r=s.\end{cases}$$

10. $x=\frac{r^{n-2}-(r-1)^{n-2}}{r^n-(r-1)^n}$ 若 $j<r$ 且 $k<r$.

$x=\frac{r^{n-2}}{r^n-(r-1)^n}$ 若 $j\leqslant r$ 和 $k=r$ 或者 $j=r$ 和 $k\leqslant r$.

$x=0$ 若 $j>r$ 或者 $k>r$.

11. $\sigma^2\approx\frac{nN^2}{(n+1)^2(n+2)}$.

12. $\mathbf{P}\{\mathbf{N}=n,\ \mathbf{K}=k\}=\binom{n}{k}p^{n-k}(qq')^k\cdot qp'$.

$\mathbf{P}\{\mathbf{N}=n\}=(1-qp')^n qp'$.

$\mathbf{P}\{\mathbf{K}=k\}=(qq')^k qp'\Sigma\binom{-k-1}{\nu}(-p)^\nu=p'q'^k$.

$$13.\ \mathbf{E}\left(\frac{\mathbf{K}}{\mathbf{N}+1}\right)=\Sigma kp_{k,n}/(n+1)=q^2p'q'\sum_{n=1}^{\infty}\left(1-\frac{1}{n+1}\right)(p+qq')^{n-1}$$

$$=\frac{qq'}{1-qp'}-\frac{p^2p'q'}{(1-qp')^2}\log\frac{1}{qp'}.$$

$$\mathbf{E}(\mathbf{K})=\frac{q'}{p'};\ \mathbf{E}(\mathbf{N})=\frac{(1-qp')}{qp'};$$

$$\mathrm{Cov}(\mathbf{K},\mathbf{N})=\frac{q}{qp'^2}.\quad \rho(\mathbf{K},\mathbf{N})=\left\{\frac{q'}{(1-qp')}\right\}^2.$$

14. $p_k=p^kq+q^kp$; $\mathbf{E}(\mathbf{X})=pq^{-1}+qp^{-1}$; $\mathrm{Var}(\mathbf{X})=pq^{-2}+qp^{-2}-2$.

15. $q_k=p^2q^{k-1}+q^2p^{k-1}$; $\mathbf{P}\{\mathbf{X}=m,\ \mathbf{Y}=n\}=p^{m+1}q^n+q^{m+1}p^n$，其中 $m,n\geqslant1$; $\mathbf{E}(\mathbf{Y})=2$; $\sigma^2=2(pq^{-1}+qp^{-1}-1)$.

17. $\binom{n}{k}364^{n-k}365^{1-n}$.

18. (a) $365\{1-364^n\cdot365^{-n}-n364^{n-1}\cdot365^{-n}\}$; (b) $n\geqslant28$.

19. (a) $\mu=n$, $\sigma^2=(n-1)n$; (b) $\mu=(n+1)/2$, $\sigma^2=(n^2-1)/12$.

20. $\mathbf{E}(\mathbf{X})=np_1$; $\mathrm{Var}(\mathbf{X})=np_1(1-p_1)$; $\mathrm{Cov}(\mathbf{X},\mathbf{Y})=-np_1p_2$.

21. $-n/36$。这是問題 20 的一个特别情形。

25. $\mathbf{E}(\mathbf{Y}_r)=\sum_{k=1}^{r}\frac{N}{r-k+1}$; $\operatorname{Var}(\mathbf{Y}_r)=\sum_{k=1}^{r}\frac{N(N-r+k-1)}{(r-k+1)^2}$.

26. (a) $1-q^k$; (b) $\mathbf{E}(\mathbf{X})=N\left\{1-q^k+\frac{1}{k}\right\}$; (c) $\frac{d\mathbf{E}(\mathbf{X})}{dk}=0$.

27. $\Sigma(1-p_j)^n$. 当第 j 类不出現时令 $\mathbf{X}_j=1$ 否則为 0.

28. $\mathbf{E}(\mathbf{X})=\frac{r_1(r_2+1)}{r_1+r_2}$; $\operatorname{Var}(\mathbf{X})=\frac{r_1r_2(r_1-1)(r_2+1)}{(r_1+r_2-1)(r_1+r_2)^2}$.

30. $\mathbf{E}(\mathbf{S}_n)=\frac{nb}{b+r}$; $\operatorname{Var}(\mathbf{S}_n)=\frac{nbr\{b+r+nc\}}{(b+r)^2(b+r+c)}$.

33. $$\mathbf{E}\left(\frac{r}{\mathbf{X}}\right)=r\sum_{k=r}^{\infty}k^{-1}\binom{k-1}{r-1}p^rq^{k-r}.$$
$$=\sum_{k=1}^{r-1}(-1)^{k-1}\frac{r}{r-k}\left(\frac{p}{q}\right)^k+\left(\frac{-p}{q}\right)^r r\log p.$$

为了由第一公式推出最后一个公式来，先令 $f(q)=r\Sigma k^{-1}\binom{k-1}{r-1}q^k$. 应用第二章(12.4)，我們发現 $f'(q)=rq^{r-1}(1-q)^{-r}$，則由重复利用分部积分法可以推出我們的結論.

参 考 文 献

[1] B. O. Koopman, The axioms and algebra of intuitive probability, *Annals of Mathematics* (2), **41** (1940), 269—292.
B. O. Koopman, The bases of probability, *Bulletin of the American Mathematical Society*, **46** (1940), 763—774.

[2] Von Mises, Wahrscheinlichkeitsrechnung, Leipzig and Wien, 1931.

[3] А. Н. Колмогоров, Основные понятия теории вероятностей, *ОНТИ*, 1936 (丁寿田譯，概率論基本概念).

[4] W. A. Whitworth, Choice and chance, 5th ed., London, 1901, reprinted by G. E. Stechert, New York, 1942.

[5] W. A. Whitworth, DCC Exercises, reprinted New York, 1945.

[6] Tables of probability functions, **1**, National Bureau of Standards (1941).

[7] Intermédiaire des recherches mathématiques, **2** (1946), 112.

[8] R. von Mises, Ueber Aufteilungs- und Besetzungs-Wahrscheinlichkeiten, *Revue de la Faculté des Sciences de l'Université d'Istanbul*, N. S., 4 (1938—1939), 145—163.

[9] H. Margenau and G. M. Murphy, The mathematics of physics and chemistry, New York, 1943, Chapter 12.

[10] F. S. Swed and C. Eisenhart, Tables for testing randomness of grouping in a sequence of alternatives, *Annals of Mathematical Statistics*, **14** (1943), 66—87.

[11] W. G. Cochran, An extension of Gold's method of examining the apparent persistence of one type of weather, *Quarterly Journal of the Royal Meteorological Society*, **64**:277 (1938), 631—634.

[12] A. Wald and J. Wolfowitz, On a test whether two samples are from the same population, *Annals of Mathematical Statistics*, **2** (1940), 147—162.

[13] N. T. J. Bailey, On estimating the size of mobile populations from recapture data, *Biometrika*, **38** (1951), 293—306.

[14] D. G. Chapman, Some properties of the hypergeometric distribution with applications to zoological sample censuses, *University of California Publications in Statistics*, **1** (1951), 131—160.

[15] James Stirling, Methodus differentialis, 1730.

[16] H. Robbins, A remark on Stirling's formula, *American Mathematical Monthly*, **62** (1955), 26—29.

[17] D. G. Catcheside, The effect of X-ray dosage upon the frequency of induced structural changes in the chromosomes of Drosophila Melanogaster, *Journal of Genetics*, **36** (1938), 307—320.

[18] J. A. Greenwood and E. E. Stuart, Review of Dr. Feller's critique, *Journal for Parapsychology*, **4** (1940), 298—319, 特别可参閲 306 頁.

[19] R. E. Greenwood, Coupon collector's test for random digits, *Mathematical Tables and Other Aids to Computation*, **9** (1955), 1—5.

[20] *Mathematical Scandinavica*, **1** (1953), 263—285; **2** (1954), 195—223.

[21] A. Dvoretzky and T. Motzkin, A problem of arrangements, *Duke Mathematical Journal*, **14** (1947), 305—313.

[22] M. T. L. Bizley, Derivation of a new formula for the number of minimal lattice paths, etc., *The Journal of the Institute of Actuaries*, **80**, Part 1. No. 354 (1954), 55—62.

[23] Galton's rank-order test, *Biometrika*, **42** (1955), 261—262.

[24] K. L. Chung and W. Feller, Fluctuations in coin tossing, *Proceedings National Academy of Sciences USA*, **35** (1949), 605—608; [23].

[25] Paul Lévy, Sur certains processus stochastiques homogènes, *Compositio Mathematica*, **7** (1939), 283—339.

[26] P. Erdös and M. Kac, On the number of positive sums of independent random variables, Bulletin of the American Mathematical Society, **53** (1947), 1011—1020.

[27] M. Fréchet, Les probabilités associées à un système d'événements compatibles et dépendants, *Actualités scientifiques et industrielles*, nos. 859 and 942, Paris, 1940 and 1943.

[28] Von Mises, Über Aufteilungs- und Besetzungswahrscheinlichkeiten, *Revue de la Faculté des Sciences de l'Université d'Istanbul*, N. S., **4** (1939), 145—163.

[29] D. G. Catcheside, D. E. Lea, and J. M. Thoday, Types of chromosome structural change introduced by the irradiation of tradescantia microspores, *Journal of Genetics*, **47** (1945—1946), 113—149.

[30] F. Eggenberger and G. Polya, Über die Statistik verketteter Vorgänge, *Zeitschrift für Angewandte Mathematik und Mechanik*, **3** (1923), 279—289.

[31] B. Friedman, A Simple urn model, *Communications on Pure and Applied Mathematics*, **2** (1949), 59—70.

[32] P. and T. Ehrenfest, Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem, *Physikalische Zeitschrift*, **8** (1907), 311—314.

[33] M. Kac, Random walk and the theory of Brownian Motion, *American Mathematical Monthly*, **54** (1947), 369—391.

[34] G. H. Hardy, Mendelian proportions in a mixed population, Letter to the Editor, *Science*, N. S., **28** (1908), 49—50.

[35] Théorie analytique des associations biologiques II, *Actualités scientifiques et industrielles*, no. 780, Paris, 1939.

[36] K. Marke, Die Gleichförmigkeit in der Welt, Munich, 1916.

[37] National Bureau of Standards, Tables of the binomial probability distribution, *Applied Mathematics Series*, **6** (1950).

[38] H. C. Romig, 50—100 Binomial tables, John Wiley and Sons, 1953.

[39] Tables of the cumulative binomial probability distribution, Harvard Computation Laboratory, 1955; Tables of the cumulative binomial probabilities, Ordnance Corps, ORDP 20-11 (1952).

[40] R. A. Fisher, Statistical Methods for research workers, Edinburgh-London, 1932, 66;

T. C. Fry, Probability and it's engineering uses, New York, 1928, 303 ff.

[41] P. V. Sukhatme and V. G. Panse, Size of experiments for testing sera or vaccines, *Indian Journal of Veterinary Science and Animal Husbandry*, **13** (1943), 75—82.

[42] Siméon, D. Poisson, Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précedées des règles générales du calcul des probabilités, 1837.

[43] M. G. Kendall and Babington Smith, Tables of random sampling numbers, *Tracts for Computers*, No. 24, Cambridge, 1940.

[44] E. C. Molina, Poisson's exponential binomial limit, New York, 1942.

[45] E. J. Gumbel, Les centenaires, *Aktuárske* Vedy, Prague, **7** (1937), 1—8.

[46] R. D. Clarke, An application of the Poisson distribution, *Journal of the Institute of Actuaries*, **72** (1946), 48.

[47] F. Thorndike, Applications of Poisson's probability summation, *The Bell System Technical Journal*, **5** (1926), 604—624.

[48] J. Neyman, Lectures and conferences on mathematical statistics (mimeographed), Dept. of Agriculture, Washington, 1938.

[49] P. A. P. Moran, A mathematical theory of animal trapping, *Biometrika*, 38 (1951), 307—311.

[50] A. Wald, Sequential tests of statistical hypotheses, *Annals of Mathematical Statistics*, **16** (1945), 166.

[51] Tables of probability functions, **2**, National Bureau of Standards, New York, 1942.

[52] Abraham DeMoivre, The doctrine of chance, 1718.

[53] Pierre, S. Laplace, Théorie analytique des probabilités, 1812.

[54] W. Feller, On the normal approximation to the binomial distribution, *Annals of Mathematical Statistics*, **16** (1945), 319—329.

[55] E. C. Molina, Probability in engineering, *Electrical Engineering*, **54** (1935), 423—427.

Bell Telephone System Technical Publications Monograph, B-854.

[56] A. Wald and J. Wolfowitz, On a test whether two samples are from the same population, Annals of Mathematical Statistics, **11** (1940), 147—162; A. M. Mood, The distribution theory of runs, *ibid.*, 367—392.

[57] N. Smirnov, Über Wahrscheinlichkeiten grosser Abweichungen (以俄文写的，德文摘要), Recueil Mathématique [Sbornik] Moscou, **40** (1933), 443—454.

[58] A. Khintchine, Über einen neuen Grenzwertsatz der Wahrscheinlichkeitsrechnung, *Mathematische Annalen*, **101** (1929), 745—752.

[59] J. L. Doob, Note on probability, *Annals of Mathematics*, **37** (1936), 363—367.

[60] A. Khintchine, Über einen Satz der Wahrscheinlichkeitsrechnung, *Fundamenta Mathematicae*, **6** (1924), 9—20.

[61] A. Kolmogoroff, Das Gesetz des iterierten Logarithmus, *Mathematische Annalen*, **101** (1929), 126—135.

[62] F. Hausdorff, Grundzüge der Mengenlehre, Leipzig, 1913.

[63] Hardy and Littlewood, Some problems of Diophantine approximation, *Acta Mathematica*, **37** (1914), 155—239.

[64] W. Feller, The general form of the so-called law of the iterated logarithm, *Transactions of the American Mathematical Society*, **54** (1943), 373—402.

[65] P. Erdös, On the law of the iterated logarithm, *Annals of Mathematics* (2), **43** (1942), 419—436.

[66] G. Polya, Eine Wahrscheinlichkeitsaufgaba zur Kundenwerbung, *Zeitschrift für Angewandte Mathematik und Mechanik*, **10** (1930), 96—97.

[67] W. Hoeffding, On the distribution of the number of successes in independent trials, *Annals of Mathematical Statistics*, **27** (1956), 713—721.

[68] A. Kolmogorov, Über die Summen zufälliger Grössen, *Mathematische Annalen*, **99** (1928), 309—319; **102** (1929), 484—488.

[69] R. Dorfman, The Detection of defective members of large populations, *Annals of Mathematical Statistics*, **14** (1943), 436—440.

人 名 对 照 表

Banach 巴拿赫
Bayes 貝依斯
Bernoulli 伯努利
Bernstein S. 伯因斯坦
Bertrand 伯川
Beta 貝塔
Bonferrni 彭費雷尼
Boole 波尔
Borel 波雷尔
Bose 波司
Cantelli 康特立
Cantor 康托
De Moivre 德模佛
Doob J. L. 杜勃
Ehrenfest 爱伦弗斯特
Einstein 爱因斯坦
Eisenhart 埃生赫脫
Erdös P. 耳多斯
Eular 尤拉
Fermat 弗馬
Fermi-Dirac 弗米-迪拉克
Fisher 費歇
Fréchet 弗雷謝
Friedman 弗雷德曼
Gauss 高斯
Hausdorff 豪司道夫
Hodges 郝捷士
Ising 艾辛
Kac M. 卡斯
ndall 肯达尔
opman 柯泼曼
place 拉普拉斯
esgue 勒貝格
Marbe 馬伯
Maxwell-Boltzmann 馬克司威尔-波茨曼
Monfourt 孟弗梯
Moran 莫莱
Morse 莫尔斯
Newman 奈依曼
Newton 牛頓
Pascal 巴斯噶
Paul Lévy 保尔·列威
Poisson 普阿松
Polya 波利亚
Rademacher 拉特馬黑尔
Riemman 黎曼
Rutherford 罗泽福特
Schwarz 希瓦兹
Shewhart 休哈特
Smith 史密斯
Spane Andersen 斯伯·安德生
Steinhaus 史特因豪司
Stirling 史特令
Swed 史怀特
Taylor 泰劳
Von Mises 馮·米賽斯
Wald 瓦尔德
Weldon 威尔頓
Whitworth 惠瓦茨
William Feller 威廉·費勒
Wolfowitz 华尔夫維兹
Колмогоров 柯尔莫格洛夫
Марков 馬尔科夫
Хинчин 辛欽
Чебышев 切比謝夫

第十章 大数定律

1. 同分布的变量

在第七章和第八章中导出的伯努利试验的极限定理是一般极限定理的特殊情况，后者我们不能在本卷中论述. 然而，为了揭示随机变量期望值的新特征，此处我们至少将讨论大数定律的某些情况.

当我们考虑 n 次伯努利试验的成功次数 $\mathbf{S}_n$ 对于 n 的依赖性时，就可以比较清楚地看出伯努利试验与随机变量理论之间的联系. 对于每次试验，$\mathbf{S}_n$ 增加 1 或 0，故我们有

$$\mathbf{S}_n=\mathbf{X}_1+\mathbf{X}_2+\cdots+\mathbf{X}_n, \tag{1.1}$$

此处当第 k 次试验出现成功时 $\mathbf{X}_k$ 等于 1，否则 $\mathbf{X}_k$ 等于 0. 于是 $\mathbf{S}_n$ 是 n 个相互独立的随机变量之和，其中每个变量分别以概率 p 与 q 取值 1 与 0. 一个直接的推广是考虑形如 (1.1) 的和，其中 $\mathbf{X}_k$ 是具有同一任意分布的独立随机变量. 第六章第 4 节的(弱)大数定律所陈述的是，当 n 很大时，平均数 $\mathbf{S}_n/n$ 多半是接近 p 的. 这是下述定律的特殊情况：

大数定律. 设 $\{\mathbf{X}_k\}$ 是独立同分布的随机变量序列. 如果期望值 $\mu=\mathbf{E}(\mathbf{X}_k)$ 存在，则对于每个 $\varepsilon>0$，当 $n\to\infty$ 时，有

$$\mathbf{P}\left\{\left|\frac{\mathbf{X}_1+\cdots+\mathbf{X}_n}{n}-\mu\right|>\varepsilon\right\}\to 0; \tag{1.2}$$

即，平均数 $\mathbf{S}_n/n$ 与期望值之间的偏差小于任意给定的 ε 的概率趋向于 1.

最先在上述一般情况证明这个定理的是辛钦[1]. 旧的证明必

1) 参见 [1]. 关于这个定理的应用范围，请读者参看第六章第 4 节末的注意.

须引进方差 $\mathrm{Var}(\mathbf{X}_k)$ 有限这个不必要的限制1)．而在这种情况下，我们有如下的精密得多的结果，它是伯努利试验的德模佛-拉普拉斯极限定理的推广，并称为

中心极限定理．设 $\{\mathbf{X}_k\}$ 是独立同分布的随机变量序列．假定 $\mu=\mathbf{E}(\mathbf{X}_k)$ 与 $\sigma^2=\mathrm{Var}(\mathbf{X}_k)$ 存在，并令 $\mathbf{S}_n=\mathbf{X}_1+\cdots+\mathbf{X}_n$．则对于每个固定的 β，

$$\mathbf{P}\left\{\frac{\mathbf{S}_n-n\mu}{\sigma n^{1/2}}<\beta\right\}\to\Phi(\beta), \tag{1.3}$$

此处 $\Phi(x)$ 是第七章第 1 节引进的正态分布函数．这个定理是麟德堡（Lindeberg）证明的[2]；在此之前，李雅普诺夫（Ляпунов）及其他作者曾在更强的限制条件下给出过它的证明．必须了解，这个定理仅仅是一个更一般的定理的特殊情况，后者我们推迟到第二卷中讨论．此处我们注意，(1.3) 比 (1.2) 强，这是因为它给出了偏差 $\left|\frac{1}{n}\mathbf{S}_n-\mu\right|$ 大于 $\sigma/n^{\frac{1}{2}}$ 的概率的一个估计．另一方面，大数定律 (1.2) 在随机变量 $\mathbf{X}_k$ 不具有有限方差的情况下也成立，故它比中心极限定理更一般．由于这个缘故我们将给出大数定律一种独立的证明，但我们先举例来说明这两个极限定理．

例．(a) 在独立掷一颗对称骰子的序列中，设 $\mathbf{X}_k$ 表示第 k 次投掷时出现的点数．则 $\mathbf{E}(\mathbf{X}_k)=(1+2+3+4+5+6)/6=3.5$，$\mathrm{Var}(\mathbf{X}_k)=(1^2+2^2+3^2+4^2+5^2+6^2)/6-3.5^2=\frac{35}{12}$．大数定律所陈述的是，当 n 很大时，平均点数 $\mathbf{S}_n/n$ 多半和 3.5 很接近．中心极限定理所陈述的则是，$|\mathbf{S}_n-3.5n|<\alpha\cdot(35n/12)^{\frac{1}{2}}$ 的概率大约是 $\Phi(\alpha)-\Phi(-\alpha)$．当 $n=1000$，$\alpha=1$ 时我们得到，$3450<\mathbf{S}_n<3550$ 的概率大约是 0.68．令 α 取中位值2) $\alpha=0.6744$，我们得到，$\mathbf{S}_n$ 位于区间 3500 ± 36 之内和位于此区间之

---

1) 马尔科夫（A. A. Марков）证明，对于某个 $a>0$，$\mathbf{E}(|\mathbf{X}_k|^{1+a})$ 存在就够了．

2) 我们称使得 $|\mathbf{S}_n-3.5|<\alpha\cdot(35n/12)^{1/2}$ 的概率等于 0.5 的 α 为中位值．此时有 $\Phi(\alpha)-\Phi(-\alpha)=0.5$，$\alpha=0.6744$，$\alpha\cdot(35n/12)^{1/2}=36.43$ ——译者注．

外的机会大致相等.

(b) **抽样.** 设在N个家庭的总体中有 N_k 个家庭恰有 k 个小孩 $\left(k=0, 1, \cdots; \sum_k N_k = N\right)$. 任选的一个家庭中的小孩数是一随机变量,它以概率 $p_\nu = N_\nu/N$ 取值 ν. 一个大小为 n 的有放回抽样的样本表示 n 个独立随机变量或 n 次"观察" $\mathbf{X}_1, \cdots, \mathbf{X}_n$, 其中每一个都具有同一分布; $\mathbf{S}_n/n$ 为样本平均值. 大数定律告诉我们,对于充分大的随机样本,样本平均值多半和 $\mu = \sum \nu p_\nu = \sum \nu N_\nu/N$ 很接近,后者正是总体平均值. 中心极限定理可以用来估计偏差的概率大小和确定为使估计可靠所必需的样本大小. 实际上 μ 和 σ^2 都是未知的. 然而, 通常容易对 σ^2 作出初步的估计, 并且总是可以作出比较安全的估计. 如果我们希望样本平均值 $\mathbf{S}_n/n$ 与未知的总体平均值之间的偏差小于 1/10 的概率为 0.99 或更大,则样本大小应该满足条件

$$\mathbf{P}\left\{\left|\frac{\mathbf{S}_n - n\mu}{n}\right| < \frac{1}{10}\right\} \geqslant 0.99. \tag{1.4}$$

$\Phi(x) - \Phi(-x) = 0.99$ 的根为 $x = 2.57\cdots$, 故 n 应当满足 $n^{\frac{1}{2}}/10\sigma \geqslant 2.57$ 或 $n \geqslant 660\,\sigma^2$. 于是所需的样本大小大致可根据对 σ^2 所作的谨慎的初步估计来确定. 类似的情况经常出现. 例如,在实验者取 n 次测量的平均值时, 他也信赖大数定律, 并用样本平均值作为未知的理论期望值的一个估计. 估计的可靠性完全由 σ^2 来确定,而通常我们不得不使用 σ^2 的相当粗糙的估值.

(c) **普阿松分布.** 在第七章第 4 节中我们曾证明, 当 λ 很大时, 普阿松分布 $\{p(k; \lambda)\}$ 能用正态分布来近似表示. 这实际上是中心极限定理的一个直接推论. 设变量 $\mathbf{X}_k$ 具有普阿松分布 $\{p(k; \gamma)\}$. 则 $\mathbf{S}_n$ 具有均值与方差为 $n\gamma$ 的普阿松分布 $\{p(k; n\gamma)\}$. 令 $\lambda = n\gamma$, 则当 $n \to \infty$ 时我们有

$$\sum_{k<\lambda+\beta\lambda^{\frac{1}{2}}} e^{-\lambda}\lambda^k/k! \to \Phi(\beta), \tag{1.5}$$

其中的和是对所有小于 $\lambda + \beta\lambda^{\frac{1}{2}}$ 的 k 来求的. 显然,当 λ 按任意

方式趋向 ∞ 时，(1.5) 仍成立. 这是一个具有广泛意义的定理，它在发散级数的求和理论中具有应用. 在估计 (1.5) 两边的差时可以采用普通理论中的方法.

关于无期望值的变量的附注

如果期望值 μ 不存在，则大数定律和中心极限定理都没有意义，但可用更为一般的定理来代替它们. 这些定理可提供同一类的讯息. 在近代理论中，无期望值的变量起着重要作用，物理中的等待时间和循环时间就属于这种类型. 甚至对于简单的掷钱币的游戏也是如此.

将 n 个钱币一一投掷. 设 $\mathbf{X}_k$ 表示对于第 k 个钱币直到正面和反面的累积数目第一次达到平衡的等待时间. 则 $\mathbf{X}_k$ 是独立同分布的随机变量：每个 $\mathbf{X}_k$ 仅取正偶数值且 $P(\mathbf{X}_k = 2r) = f_{2r}$，此处概率分布 $\{f_{2r}\}$ 按第三章 (4.2) 定义. 根据第三章第 4 节的定理 3，和 $\mathbf{S}_n = \mathbf{X}_1 + \cdots + \mathbf{X}_n$ 的分布由下式给出：

$$\mathbf{P}(\mathbf{S}_n = 2r) = f_{2r}^{(n)} \tag{1.6}$$

其中 $f_{2r}^{(n)}$ 按第三章 (4.11) 定义. 在第三章第 8 节 (c) 中我们曾证明，当 $n \to \infty$ 时，

$$\mathbf{P}\{\mathbf{S}_n < n^2x\} \to 2[1 - \Phi(x^{-\frac{1}{2}})]. \tag{1.7}$$

这是一个与中心极限定理具有相同特征的极限定理，两者的显著差别是此处是变量 $\mathbf{S}_n/n^2$ 而不是 $\mathbf{S}_n/n$ 具有极限分布.

用物理语言来说，$\mathbf{X}_k$ 表示对同一量的独立测量，而这极限定理则断言，在概率意义下，平均值 $\mathbf{S}_n/n$ 与 n 成比例地增加. 这个性质的出人意料的推论我们已在第三章中讨论过[1].

*2. 大数定律的证明

我们分两步来进行. 首先假定 $\sigma^2 = \mathrm{Var}(\mathbf{X}_k)$ 存在. 根据第

1) 对于不具有有限期望值的变量，与大数定律类似的定理可参看第 4 节及问题13.

* 本节讨论专门的论题，初读时可以略去.

九章(5.6)中的加法规则可得 $\text{Var}(\mathbf{S}_n) = n\sigma^2$. 根据第九章中的切比雪夫不等式(6.1), 对于每个 $t > 0$, 我们有

$$\mathbf{P}\{|\mathbf{S}_n - n\mu| > t\} \leqslant \frac{n\sigma^2}{t^2}. \tag{2.1}$$

令 $t = \varepsilon n$[1], 则上式左边小于或等于 $\sigma^2/\varepsilon^2 n$, 而当 $n \to \infty$ 时, $\sigma^2/\varepsilon^2 n$ 趋向于零. 于是定律得证.

其次我们放弃 $\text{Var}(\mathbf{X}_k)$ 存在的限制. 这种情况可用截尾法化为前面所讨论的情况. 这方法是一种重要的标准方法. 定义两组依赖于 $\mathbf{X}_k$ 的随机变量如下:

$$\begin{aligned} &\mathbf{U}_k = \mathbf{X}_k, \quad \mathbf{V}_k = 0, \quad \text{如果 } |\mathbf{X}_k| \leqslant \varepsilon n; \\ &\mathbf{U}_k = 0, \quad \mathbf{V}_k = \mathbf{X}_k, \quad \text{如果 } |\mathbf{X}_k| > \varepsilon n. \end{aligned} \tag{2.2}$$

此处 $k = 1, 2, \cdots$; n 与 ε 是固定的. 于是恒有

$$\mathbf{X}_k = \mathbf{U}_k + \mathbf{V}_k. \tag{2.3}$$

设 $\{f(x_j)\}$ 是变量 $\mathbf{X}_k$ 的共同概率分布. 由于我们假定了 $\mu = \mathbf{E}(\mathbf{X}_k)$ 存在, 故和

$$\sum |x_j| f(x_j) = A \tag{2.4}$$

存在. 现在

$$\mu_n' = \mathbf{E}(\mathbf{U}_k) = \sum_{|x_j| \leqslant \varepsilon n} x_j f(x_j), \tag{2.5}$$

此处的和是对一切使得 $|x_j| \leqslant \varepsilon n$ 成立的 j 来求的. 显然当 $n \to \infty$ 时, $\mu_n' \to \mu$, 故对于任意的 $\delta > 0$, 当 n 充分大时有

$$|\mu_n' - \mu| < \delta. \tag{2.6}$$

又由(2.5)与(2.4)有

$$\text{Var}(\mathbf{U}_k) \leqslant \mathbf{E}(\mathbf{U}_k^2) \leqslant \varepsilon n \sum_{|x_j| \leqslant \varepsilon n} |x_j| f(x_j) \leqslant \varepsilon A n. \tag{2.7}$$

$\mathbf{U}_k$ 是相互独立的, 且其和 $\mathbf{U}_1 + \mathbf{U}_2 + \cdots + \mathbf{U}_n$ 完全可用前面在方差有限的情况讨论 $\mathbf{X}_k$ 的方法来讨论; 应用切比雪夫不等式,

---

1) 原书误为 $t > \varepsilon n$ ——译者注.

我们得到如下的类似于(2.1)的不等式:

$$\mathbf{P}\left\{\left|\frac{\mathbf{U}_1+\cdots+\mathbf{U}_n}{n}-\mu_n'\right|>\delta\right\}\leqslant\frac{\operatorname{Var}(\mathbf{U}_k)}{n\delta^2}\leqslant\frac{\varepsilon A}{\delta^2}. \quad (2.8)$$

由上式及(2.6)有

$$\mathbf{P}\left\{\left|\frac{\mathbf{U}_1+\cdots+\mathbf{U}_n}{n}-\mu\right|>2\delta\right\}\leqslant\frac{\varepsilon A}{\delta^2}. \quad (2.9)$$

易见，$\mathbf{V}_k=0$ 的概率很大. 事实上

$$\mathbf{P}\{\mathbf{V}_k\neq 0\}=\sum_{|x_j|>\varepsilon n}f(x_j)\leqslant\frac{1}{\varepsilon n}\sum_{|x_j|>\varepsilon n}|x_j|f(x_j), \quad (2.10)$$

且当 n 增加时上式右端的和趋向于 0. 故当 n 充分大时,有

$$\mathbf{P}\{\mathbf{V}_k\neq 0\}\leqslant\frac{\varepsilon}{n}, \quad (2.11)$$

于是由第一章的基本不等式(7.6)有

$$\mathbf{P}\{\mathbf{V}_1+\cdots+\mathbf{V}_n\neq 0\}\leqslant\varepsilon. \quad (2.12)$$

现在 $\mathbf{S}_n=(\mathbf{U}_1+\cdots+\mathbf{U}_n)+(\mathbf{V}_1+\cdots+\mathbf{V}_n)$，故由(2.9)与(2.12)有

$$\mathbf{P}\left\{\left|\frac{\mathbf{S}_n}{n}-\mu\right|>2\delta\right\}\leqslant\mathbf{P}\left\{\left|\frac{\mathbf{U}_1+\cdots+\mathbf{U}_n}{n}-\mu\right|>2\delta\right\}+$$

$$+\mathbf{P}\{\mathbf{V}_1+\cdots+\mathbf{V}_n\neq 0\}\leqslant\frac{\varepsilon A}{\delta^2}+\varepsilon. \quad (2.13)$$

因为 ε 与 δ 是任意的,故可使上式右边任意小,于是定理得证.

3. “公平”赌博理论

为了进一步分析大数定律的含意，我们将使用由来已久的赌博术语,但此处的讨论同样可以应用到实用性较强的问题中去,并且我们的两个基本假设在统计学和物理学中的意义比它们在赌场中的意义更为现实. 首先,我们假定赌徒具有无限的赌本,因而不论输掉多少他仍能继续赌博.（如果没有这个假设,则要考虑赌徒输光的问题——概率论的研究者早就对它感到兴趣. 在瓦尔德的

序贯分析和在随机过程中这个问题是重要的，我们将在第十四章中讨论.）其次，我们假定赌徒不能随意终止赌博；试验的次数必须事先确定而不依赖于赌博的进程．实际上，一个拥有无限赌本的赌徒会等待一个好运气并在恰当的时机退出赌博．他所感兴趣的不是指定时刻的或然状态，而仅仅是在长期赌博中的那些最大的起伏．为阐明这个问题，与其用大数定律倒不如用迭对数定律（参看第八章第5节）.

设有一个赌徒，一直参加同一种赌博，$\mathbf{X}_k$ 表示他在第 k 次试验中的(正的或负的)赢利．则和 $\mathbf{S}_n=\mathbf{X}_1+\cdots+\mathbf{X}_n$ 表示在 n 次独立试验中的累积赢利．如果对于每次试验赌徒需付入场费 μ'（不一定是正的)，则 $n\mu'$ 表示累积入场费，而 $\mathbf{S}_n-n\mu'$ 则表示净累积赢利．当 $\mu=\mathbf{E}(\mathbf{X}_k)$ 存在时可以应用大数定律．粗略地地说，当 n 很大时，差 $\mathbf{S}_n-n\mu$ 多半要比 n 小的多，故如果入场费 μ' 小于 μ，则当 n 很大时，赌徒的净赢利多半是正的且其大小的阶为 $n(\mu-\mu')$．根据同样的理由，如果 $\mu'>\mu$，则赌徒会要吃亏．简言之，$\mu'<\mu$ 的情况对赌徒有利，$\mu'>\mu$ 的情况对赌徒不利.

注意，以上的讨论并未涉及 $\mu'=\mu$ 的情况．在这种情况，仅能作出如下的论断：当 n 充分大时，累积赢利或损失 $\mathbf{S}_n-n\mu$ 的绝对值[1]比 n 要小得多的概率接近于1．但并未说 $\mathbf{S}_n-n\mu$ 是多半为正还是多半为负，亦即是对赌徒有利还是不利．在古典理论中称 $\mu'=\mu$ 为“公平”价格，并称 $\mu'=\mu$ 的赌博为“公平”赌博，是考虑不周密的．这个不正确的名称造成了很多危害．必须指出，所谓“公平”赌博，既可能对赌徒有利也可能对赌徒不利.

显然，在“正常”情况下，$\mathbf{E}(\mathbf{X}_k)$ 与 $\mathrm{Var}(\mathbf{X}_k)$ 都存在．对于这种情况，大数定律可由中心极限定理加以补充，后者告诉我们，对于“公平”赌博，净累积赢利 $\mathbf{S}_n-n\mu$ 的大小大致与 $n^{\frac{1}{2}}$ 同阶，且当 n 很大时，净赢利为正或为负的可能性差不多相等．故当可

1）原文无“绝对值”三字，似不妥——译者注.

以应用中心极限定理时，“公平”这一术语是合理的，但甚至在这种情况，我们仍然是与涉及长时间过程的极限定理打交道.

例如，设某赌徒参加一种赌博. 他每次以 10^{-6} 的概率净赢 10^6-1 元，而以 $1-10^{-6}$ 的概率损失 $\mu'=1$（元）的入场费. 可以把这种赌博看成是伯努利试验，易知这是一种“公平”赌博. 在一百万次试验中赌徒需付一百万元的入场费. 他可能赢 0, 1, 2, … 次. 由二项分布的普阿松近似公式知，赢 k 次的概率为 $e^{-1}/k!$（精确到好几位小数）. 于是赌徒损失一百万元的概率为 0.368…，仅能收回所花入场费的概率也是 0.368…，净赢一百万元的概率则为 0.184…，等等. 此处的 10^6 次试验等价于赢利按普阿松分布的赌博中的一个单独试验（这可通过两付牌的配对问题来理解；参见第四章第 4 节）. 显然在这种情形大数定律不具有操作上（operationally）的意义. 现在所有火灾、车祸及其它类似项目的保险，都属于上述这种类型；当被保险的意外事件发生时公司要支付巨额的赔偿费，但相应的概率非常小. 此外，被保险户通常每年仅参与一次试验，故试验次数 n 决不会变得很大，对于他来说赌博必然是“不公平”的，但它通常在经济上有好处；大数定律与他毫无关系. 对于保险公司来说，它进行赌博的次数则很多，但由于方差很大，故随机起伏是必然的. 保险费必须定得使任一年中的巨大损失能得以消除. 故与公司有关的与其说是大数定律倒不如说是输光问题.

如果方差无限，则“公平”赌博这一术语完全是一种误称；并没有什么理由可以认为净累积赢利 $\mathbf{S}_n-n\mu'$ 的起伏是围绕零的. 事实上，存在“公平”赌博的例子[3]，在其中赌徒不断吃亏的概率趋向于 1. 大数定律仅断言，赌徒净损失的大小大概具有比 n 更小的阶. 然而，并不能作出进一步的论断. 设 a_n 是满足条件 $a_n/n\to 0$ 的任意序列，则可以构造一种“公平”赌博，使得赌徒在第 n 次试验时净累积损失超过 a_n 的概率趋向于 1. 在问题 15 中，实际上可以确信，赌徒的损失将超过 $n/\log n$. 这个赌博是“公平”的，且入场费为 1. 很难想像，一个赌徒，在他遭受的损失实际上必然是稳

定地上升的情况下，会认为赌博是"公平"的.

*4. 彼得堡赌博

在古典理论中，期望值的概念并未明显地从概率的定义中分离出来，也没有建立论述这个概念的数学形式体系. 因此，具有无限期望值的随机变量就引起了不可克服的困难，甚至近代概率的研究者在最近的讨论中对它都感到生疏. 具有无限期望值的随机变量的重要性，我们已在第 1 节末强调过，本节我们将给出关于这种变量的一个与大数定律相类似的例子. 为此我们使用历史上著名的所谓彼得堡（Petersburg）悖论[1].

彼得堡赌博中的单独试验是投掷一枚钱币直到出现正面为止；如果正面在第 r 次投掷时出现，则赌徒赢得 2^r 元. 换句话说，每次试验中的赢利是分别以概率 $2^{-1}, 2^{-2}, 2^{-3}, \cdots$ 取值 $2^1, 2^2, 2^3, \cdots$ 的随机变量. 期望值形式地由 $\sum x_r f(x_r)$ 来定义，此处 $x_r = 2^r$，$f(x_r) = 2^{-r}$，故这个级数中的每一项均为 1. 于是赢利不具有有限的期望值，所以不能应用大数定律. 现在我们修改赌博规则如下：如果试验需要投掷 N 次以上(即接连出现 N 次反面)，则赌徒的赢利为 0. 修改后的赌博不及修改前那样对赌徒有利. 在新赌博中，赢利具有有限的期望值，且可以应用大数定律. 由此可得，对于每个 N，在 n 次试验后累积赢利多半要超过 nN. 因此不论赌徒对每次试验需付的固定入场费 μ' 是多少，他总能指望赢得净利[2]. 对于每个 μ' 都是如此，但为了有希望得到正的赢利，μ' 越大时 n 也必须越大. 古典理论断定，$\mu' = \infty$ 是"公平"的入场费，但近代的研究者简直不明白这个"悖论"的这种神秘的讨论.

---

* 标有星号的节讨论专门的问题，初读时可以略去.

1) 这个悖论曾由 D. 贝努利讨论过. 注意：贝努利试验是由于 J. 贝努利而命名的.

2) 这里的赌博是指彼得堡赌博. 为了理解此处的论断，对于任意固定的 μ' 取 $N > \mu'$，并设想赌博按上述规定修改，于是可应用大数定律——译者注.

确定入场费使得彼得堡赌博具有古典意义下的“公平”赌博的所有性质是完全可能的，只是这时入场费不是常数而是依赖于试验的次数．在赌场中并不会遇到可变入场费的情况，在那里由于资本有限，彼得堡赌博也是不可能的．在期望值 $\mu = \mathbf{E}(\mathbf{X}_k) > 0$ 为有限的情况，如果当 n 很大时累积赢利 $\mathbf{S}_n$ 与累积入场费 $e_n = n\mu'$ 之比大致接近于 1（即差 $\mathbf{S}_n - e_n$ 的阶大致要比 $e_n = n\mu'$ 的阶更小），则称赌博是“公平”的．如果 $\mathbf{E}(\mathbf{X}_k)$ 不存在，则我们不能令 $e_n = n\mu'$ 而必须用其它的方法来确定 e_n．如果对于每个 $\varepsilon > 0$

$$\mathbf{P}\left\{\left|\frac{\mathbf{S}_n}{e_n} - 1\right| > \varepsilon\right\} \to 0, \tag{4.1}$$

则我们称累积入场费为 e_n 的赌博是在古典意义下的公平赌博．(4.1) 式与大数定律完全类似，对于后者 $e_n = n\mu'$．大数定律被物理学家解释为 n 次独立测量的平均值必定接近于 μ．在现在的场合 n 次测量的平均值必定接近于 e_n/n．极限定理 (4.1)，当它成立时，具有与大数定律相同的、数学和实用上的意义．

现在我们将证明[1)]，如果令 $e_n = n\log n$，则彼得堡赌博成为古典意义下的“公平”赌博，其中 $\log n$ 是以 2 为底的对数，即 $2^{\log n} = n$．

证．我们用第 2 节中的截尾法，此时变量 $\mathbf{U}_k$ 与 $\mathbf{V}_k(k = 1, 2, \cdots, n)$ 按下式定义：

$$\begin{aligned} &\mathbf{U}_k = \mathbf{X}_k, \quad \mathbf{V}_k = 0, \quad \text{如果 } \mathbf{X}_k \leqslant n\log n; \\ &\mathbf{U}_k = 0, \quad \mathbf{V}_k = \mathbf{X}_k, \quad \text{如果 } \mathbf{X}_k > n\log n. \end{aligned} \tag{4.2}$$

又有 $\mathbf{X}_k = \mathbf{U}_k + \mathbf{V}_k$，且 $\mathbf{U}_k$ 相互独立．对于每个 t 我们有 $\mathbf{P}\{\mathbf{X}_k > t\} < 2/t$，故有 $\mathbf{P}\{\mathbf{V}_k \neq 0\} < 2/(n\log n)$，或

$$\mathbf{P}\{\mathbf{V}_1 + \mathbf{V}_2 + \cdots + \mathbf{V}_n > 0\} < \frac{2}{\log n} \to 0. \tag{4.3}$$

为证明 (4.1) 只需证明

---

1) 这是广义大数定律的一个特殊情况，根据这个定律易导出 (4.1) 成立的充分必要条件；参见[4]．

$$\mathbf{P}\{|\mathbf{U}_1+\cdots+\mathbf{U}_n-n\log n|>\varepsilon n\log n\}\to 0. \tag{4.4}$$

令 $\mu_n=\mathbf{E}(\mathbf{U}_k)$，$\sigma_n^2=\operatorname{Var}(\mathbf{U}_k)$；这些量依赖于 n，但对 $\mathbf{U}_1$，$\mathbf{U}_2$，$\cdots$，$\mathbf{U}_n$ 来说是共同的．设 r 是满足不等式 $2^r\leqslant n\log n$ 的最大整数，则 $\mu_n=r$ 且对于充分大的 n 有

$$\log n<\mu_n\leqslant\log n+\log\log n. \tag{4.5}$$

类似有

$$\sigma_n^2<\mathbf{E}(\mathbf{U}_k^2)=2+2^2+\cdots+2^r<2^{r+1}\leqslant 2n\log n. \tag{4.6}$$

因为和 $\mathbf{U}_1+\cdots+\mathbf{U}_n$ 的均值与方差分别为 $n\mu_n$ 与 $n\sigma_n^2$，故由切比雪夫不等式有

$$\mathbf{P}\{|\mathbf{U}_1+\cdots+\mathbf{U}_n-n\mu_n|>\varepsilon n\mu_n\}$$
$$\leqslant\frac{n\sigma_n^2}{\varepsilon^2n^2\mu_n^2}<\frac{2}{\varepsilon^2\log n}\to 0. \tag{4.7}$$

由 (4.5) 知 $\mu_n\sim\log n$，故 (4.7) 等价于 (4.4)．

5. 分布不同的变量

直到现在为止我们仅考虑了变量 $\mathbf{X}_k$ 具有相同分布的情况．这种情况对应于同一随机游戏的重复．现在我们来考虑游戏的类型在每一步都发生变化时将会出现什么情况．这个问题更有趣味．没有必要使用赌博的语言；一个统计学家在应用统计检验时就会遇到这类随机变量的分布随着情况的不同而变化的问题．

为确定起见，我们设想对于所给概率分布的无穷序列和每个 n，总存在 n 个具有指定分布的相互独立的随机变量．我们假定均值与方差存在并令

$$\mu_k=\mathbf{E}(\mathbf{X}_k),\quad \sigma_k=\operatorname{Var}(\mathbf{X}_k). \tag{5.1}$$

和 $\mathbf{S}_n=\mathbf{X}_1+\cdots+\mathbf{X}_n$ 也具有有限的均值与方差

$$m_n=\mathbf{E}(\mathbf{S}_n),\quad s_n^2=\operatorname{Var}(\mathbf{S}_n), \tag{5.2}$$

且

$$m_n=\mu_1+\cdots+\mu_n,\quad s_n^2=\sigma_1^2+\cdots+\sigma_n^2 \tag{5.3}$$

[参见第九章公式 (2.4) 与 (5.6)]．在同分布的特殊情况我们有

$m_n = n\mu$, $s_n^2 = n\sigma^2$.

如果对于每个 $\varepsilon > 0$ 有

$$P\left\{\frac{|\mathbf{S}_n - m_n|}{n} > \varepsilon\right\} \to 0, \tag{5.4}$$

则称对于序列 $\{\mathbf{X}_k\}$ (弱)大数定律成立.

如果对于每个固定的 $\alpha < \beta$ 有

$$P\left\{\alpha < \frac{\mathbf{S}_n - m_n}{s_n} < \beta\right\} \to \Phi(\beta) - \Phi(\alpha), \tag{5.5}$$

则称序列 $\{\mathbf{X}_k\}$ 服从中心极限定理.

概率论的一个突出的特征是大数定律和中心极限定理对十分广泛的一类序列 $\{\mathbf{X}_k\}$ 成立. 特别,当 $\mathbf{X}_k$ 一致有界,亦即当存在常数 A 使得 $|\mathbf{X}_k| < A$ 对一切 k 成立时,大数定律成立. 更一般的说,大数定律成立的一个充分条件是

$$\frac{s_n}{n} \to 0. \tag{5.6}$$

这是切比雪夫不等式的直接推论, 并可用第 2 节第一段中的方法来证明. 注意,条件 (5.6) 并不是必要的(参见问题 14).

关于中心极限定理曾经得到过各种各样的充分条件, 但这些条件都被林德堡[1] 定理所代替, 按照这个定理,中心极限定理成立的一个充分条件是,对于每个 $\varepsilon > 0$ 截尾变量 $\mathbf{U}_k$ 满足条件 $s_n \to \infty$ 及

$$\frac{1}{s_n^2}\sum_{k=1}^{n} E(\mathbf{U}_k^2) \to 1, \tag{5.7}$$

其中 $\mathbf{U}_k$ 按下式定义:

$$\begin{aligned} \mathbf{U}_k &= \mathbf{X}_k - \mu_k, \quad &\text{如果 } |\mathbf{X}_k - \mu_k| \leqslant \varepsilon s_n, \\ \mathbf{U}_k &= 0, \quad &\text{如果 } |\mathbf{X}_k - \mu_k| > \varepsilon s_n. \end{aligned} \tag{5.8}$$

如果 $\mathbf{X}_k$ 一致有界,即,如果 $|\mathbf{X}_k| < A$, 则当 n 充分大使得 $s_n > 2A\varepsilon^{-1}$ 时, $\mathbf{U}_k = \mathbf{X}_k - \mu_k$. 于是 (5.7) 的左边等于 1. 故由

1) J. W. Lindeberg, 见参考文献[2].

麟德堡定理可以推出，每个一致有界的相互独立的随机变量序列 $\{\mathbf{X}_k\}$ 都服从中心极限定理(当然,还需假定 $s_n \to \infty$)．可以证明，麟德堡条件也是 (5.5) 成立的必要条件[1]． 证明将在第二卷中给出,在那里我们还将估计 (5.5) 两边的差．

我们曾指出，在变量 $\mathbf{X}_k$ 具有相同分布的情况中心极限定理比大数定律强．在一般情况并不是这样,我们也将看到,中心极限定理可适用于并不服从大数定律的序列．

例. (a) 设 $\lambda > 0$ 为常数，令 $\mathbf{X}_k = \pm k^\lambda$，其中每个值具有概率 1/2 (例如，掷一个钱币，第 k 次投掷的赌注为 $\pm k^\lambda$)．此处 $\mu_k = 0$，$\sigma_k^2 = k^{2\lambda}$，且[2]

$$s_n^2 = 1^{2\lambda} + 2^{2\lambda} + 3^{2\lambda} + \cdots + n^{2\lambda} \sim \frac{n^{2\lambda+1}}{2\lambda+1}. \qquad (5.9)$$

如果 $\lambda < \frac{1}{2}$，则条件 (5.6) 满足．故当 $\lambda < \frac{1}{2}$ 时大数定律成立;我们接着证明当 $\lambda \geqslant \frac{1}{2}$ 时大数定律不再成立．

当 $k = 1, 2, \cdots, n$ 时我们有 $|\mathbf{X}_k| = k^\lambda \leqslant n^\lambda$，故当 $n > (2\lambda+1)\varepsilon^{-2}$ 时截尾变量 $\mathbf{U}_k$ 与 $\mathbf{X}_k$ 恒等，于是当 $\lambda > 0$ 时麟德堡条件满足,故有

$$\mathbf{P}\left\{\alpha < \left(\frac{2\lambda+1}{n^{2\lambda+1}}\right)^{\frac{1}{2}} \mathbf{S}_n < \beta\right\} \to \Phi(\beta) - \Phi(\alpha). \qquad (5.10)$$

由此可得 $\mathbf{S}_n$ 大概与 $n^{\lambda+\frac{1}{2}}$ 同阶,故当 $\lambda \geqslant \frac{1}{2}$ 时大数定律不成立．在这个例子中我们看到,对于所有的 $\lambda > 0$ 中心极限定理成立,而仅当 $\lambda < \frac{1}{2}$ 时大数定律才成立．

(b) 考虑由各自投掷钱币 1000 次所构成的两个独立序列,我

---

1) 见 [20]． 在这篇文献中还导出了对于无期望值的变量也适用的广义中心极限定理．注意，此处我们仅考虑独立变量；对于相依变量麟德堡条件既不必要也不充分．

2) 根据 Stolz 定理;参见 Фихтенгольц, 微积分学教程,一卷一分册，p. 59——译者注．

们来考察出现正面的次数之差 D. 将这两个序列中的各次投掷依次从 1 到 1000 及从 1001 到 2000 加以编号，并定义 2000 个随机变量 $\mathbf{X}_k$ 如下：如果第 k 次投掷时出现反面，则令 $\mathbf{X}_k=0$；如果出正面，则当 $k\leqslant 1000$ 时令 $\mathbf{X}_k=1$ 而当 $x>1000$ 时令 $\mathbf{X}_k=-1$. 则 $\mathbf{D}=\mathbf{X}_1+\mathbf{X}_2+\cdots+\mathbf{X}_{2000}$. 且当 $k\leqslant 1000$ 时 $\mu_k=\frac{1}{2}$，当 $k>1000$ 时 $\mu_k=-\frac{1}{2}$；$\sigma_k^2=\frac{1}{4}$，$m_{2000}=0$，$s_{2000}^2=500$. 所以 D 在范围 $\pm(500)^{\frac{1}{2}}\alpha$ 之内的概率近似地等于 $\Phi(\alpha)-\Phi(-\alpha)$，且和在 2000 次投掷中正面出现次数与其期望值 1000 之间的偏差 $\mathbf{S}_{2000}-1000$ 是同等级的.

(c) 下面我们给出中心极限定理在遗传理论中的一个应用，以显示中心极限定理的推论的广泛性. 在第五章第 5 节中我们曾研究过本质上仅依赖于一对基因(纯粹单位形质)的性状. 其它一些性状(例如体长)，可以设想为很多基因对的累积效果. 为简单起见，假定对于每对特定的基因存在三种遗传型 AA，Aa 或 aa. 设它们对体长的相应贡献为 x_1，x_2 与 x_3. 一个个体的基因型是一个随机事件，一对特定的基因对体长的贡献则是以一定的概率取值 x_1，x_2，x_3 的一个随机变量. 体长是很多这样的随机变量 $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ 的积累效果. 由于每一贡献皆甚微小，我们可以粗略假定体长为和数 $\mathbf{X}_1+\cdots\mathbf{X}_n$，并不是所有的 $\mathbf{X}_k$ 都是相互独立的. 然而，中心极限定理对于很大一类相依随机变量也成立，而且也有一定的理由把 $\mathbf{X}_k$ 中的大多数当作独立随机变量来看待. 这些考虑可以进一步精确化，此处它们仅用来说明怎样用中心极限定理来解释为什么很多生物统计性状的经验分布与正态分布很接近. 这种理论也可以对遗传性状，例如子女的平均身长对其双亲的体长的依赖性，作出预测. 这种生物统计的研究始于格尔顿(F. Galton)与皮尔逊(K. Pearson).

*6. 在组合分析中的应用

本节中，我们将给出中心极限定理在与概率论并无直接关系的问题中的两个应用. 两者都与 n 个元素 $a_1, a_2, \cdots, a_n$ 的 $n!$ 种排列有关，对于每个这样的排列我们都赋与概率 $1/n!$.

(a) 逆序. 在一个给定的排列中，我们称元素 a_k 产生 r 个逆序，如果恰有 r 个足标比它小的元素(即按自然顺序在 a_k 之前的元素)在它之后. 例如，在 $(a_3a_6a_1a_5a_2a_4)$ 中，元素 a_1 与 a_2 不产生逆序，a_3 产生两个逆序，a_4 不产生逆序，a_5 产生两个逆序，a_6 产生四个逆序. 在 $(a_6a_5a_4a_3a_2a_1)$ 中，元素 a_k 产生 $k-1$ 个逆序，逆序的总数为 15. a_k 所产生的逆序数 $\mathbf{X}_k$ 是一随机变量，而 $\mathbf{S}_n = \mathbf{X}_1 + \cdots + \mathbf{X}_n$ 则是逆序的总数. 这里 $\mathbf{X}_k$ 取值 $0, 1, \cdots, k-1$ 的概率都是 $1/k$，所以

$$\mu_k = \frac{k-1}{2}, \tag{6.1}$$

$$\sigma_k^2 = \frac{1 + 2^2 + \cdots + (k-1)^2}{k} - \left(\frac{k-1}{2}\right)^2 = \frac{k^2-1}{12}.$$

a_k 所产生的逆序的数目并不依赖于 $a_1, a_2, \cdots, a_{k-1}$ 的相对次序，且 $\mathbf{X}_k$ 是相互独立的. 由 (6.1) 我们得到

$$m_n = \frac{1 + 2 + \cdots + (n-1)}{2} = \frac{n(n-1)}{4} \sim \frac{n^2}{4} \tag{6.2}$$

及

$$s_n^2 = \frac{1}{12}\sum_{k=1}^{n}(k^2-1) = \frac{2n^3 + 3n^2 + 5n}{72} \sim \frac{n^3}{36}. \tag{6.3}$$

当 n 很大时我们有 $\varepsilon s_n > n \geqslant |\mathbf{X}_k - \mu_k|$，故林德堡条件中的变量 $\mathbf{U}_k$ 与 $\mathbf{X}_k - \mu_k$ 恒等. 于是中心极限定理成立，由此我们得到如下结论：其逆序数在范围 $\frac{n^2}{4} \pm \frac{\alpha}{6} n^{\frac{3}{2}}$ 之内的排列的数目渐近

---

* 本节讨论特殊的论题，可以略去.

地由 $n!\{\Phi(\alpha)-\Phi(-\alpha)\}$ 给出．特别，在所有的排列中大约有一半其逆序的数目在范围 $(n^2/4)\pm(0.11)n^{\frac{3}{2}}$ 之内．

(b) 循环．每个排列都可以分解为若干个循环，所谓循环是指在其内部进行排列的一组元素[1]．例如，在 $(a_3a_6a_1a_5a_2a_4)$ 中，我们看到，a_1 与 a_3 互相交换，剩下的四个元素则在其内部进行排列；故这个排列包含两个循环．如果一元素排在它的自然位置，则它构成一个循环，故按自然顺序的排列 $(a_1, a_2, \cdots, a_n)$ 包含 n 个循环．另一方面，形如 $(a_2, a_3, \cdots, a_n, a_1)$，$(a_3, a_4, \cdots, a_n, a_1, a_2)$ 的循环排列都只包含一个循环．为了研究循环，用标志元素所占据的位置的箭头来表示排列是方便的．例如，$1\to3\to4\to1$ 表示 a_1 在第三个位置，a_3 在第四个位置，a_4 在第一个位置，这个循环在第三步完成．继续用这种方法来表示包含 a_2 的循环，此处 a_2 是按自然顺序的第二个元素．按照这种记号，排列 $(a_4, a_8, a_1, a_3, a_2, a_5, a_7, a_6)$ 可以表示为：$1\to3\to4\to1$；$2\to5\to6\to8\to2$；$7\to7$．

如果在此种结构中第 k 步完成一个循环，则令 $\mathbf{X}_k=1$；否则令 $\mathbf{X}_k=0$．（在最后的一个例子中，$\mathbf{X}_3=\mathbf{X}_7=\mathbf{X}_8=1$，$\mathbf{X}_1=\mathbf{X}_2=\mathbf{X}_4=\mathbf{X}_5=\mathbf{X}_6=0$．）显然，当且仅当 a_1 在第一个位置时，$\mathbf{X}_1=1$．在第 $1, 2, \cdots, n$ 步，我们分别有 $n, n-1, \cdots, 0$ 种选择方法，在其中恰有一种方法导致一个循环的完成．所以[2] $\mathbf{X}_k=1$ 的概率为 $1/(n-k+1)$，$\mathbf{X}_k=0$ 的概率为 $(n-k)/(n-k+1)$．变量 $\mathbf{X}_k$ 是相互独立的，其均值与方差为

$$\mu_k=\frac{1}{n-k+1},\quad \sigma_k^2=\frac{n-k}{(n-k+1)^2},\tag{6.4}$$

由此

---

1) 此处的叙述过于简略，今补充如下：在一个排列中，设 a_{i_1} 排在第 i_2 个位置，a_{i_2} 排在第 i_3 个位置，…，$a_{i_{k-1}}$ 排在第 i_k 个位置，而 a_{i_k} 则排在第 i_1 个位置，则称这 k 个元素构成一个循环．有关的讨论可参看张禾瑞，《近世代数基础》，pp. 54—57 ——译者注．

2) 形式上，$\mathbf{X}_k$ 的分布不仅依赖于 k 而且也依赖于 n．按从 $k=n$ 开始到 $k=1$ 为止的次序将 $\mathbf{X}_k$ 重新编号，就可使得分布仅依赖于足标．

$$m_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \sim \log n, \qquad (6.5)$$

且

$$s_n^2 = \sum_{k=1}^{n} \frac{n-k}{(n-k+1)^2} \sim \log n. \qquad (6.6)$$

$\mathbf{S}_n = \mathbf{X}_1 + \cdots + \mathbf{X}_n$ 是循环的总数. 其平均值为 m_n; 循环数在 $\log n + \alpha(\log n)^{\frac{1}{2}}$ 与 $\log n + \beta(\log n)^{\frac{1}{2}}$ 之间的排列数渐近地由 $n!\{\Phi(\beta) - \Phi(\alpha)\}$ 给出. 改进后的中心极限定理给出的估计更为精确[1].

*7. 强大数定律

(弱)大数定律(5.4)断言, 对于各个充分大的 n 偏差 $|\mathbf{S}_n - m_n|$ 大致上要比 n 小得多. 关于伯努利试验(第八章), 我们曾指出,由大数定律并不能得出对于所有充分大的 n, $|\mathbf{S}_n - m_n|/n$ 始终很小的结论;可能遇到这种情况: 大数定律成立而 $|\mathbf{S}_n - m_n|/n$ 一直在有限或无限的范围内摆动. 由大数定律仅能得出这样的结论: $|\mathbf{S}_n - m_n|/n$ 取大的值的情况是很少出现的.

我们说序列 $\mathbf{X}_k$ 服从强大数定律, 如果对于每个 $\varepsilon > 0$ 与 $\delta > 0$,有 N 与之对应,使得对于每个 $r > 0$ 所有 r 个不等式

$$\frac{|\mathbf{S}_n - m_n|}{n} < \varepsilon, \quad n = N, N+1, \cdots, N+r \qquad (7.1)$$

同时成立的概率不小于 $1 - \delta$.

(7.1)可以粗略地解释如下: 对于所有的 $n > N$, $|\mathbf{S}_n - m_n|/n$ 始终很小[2]的概率占压倒的优势.

---

1) 组合分析中的一大批渐近估计是用其它方法导出的, 见 [6]. 现在的方法较简单,但应用的范围较小, 参看 [7].

* 本节讨论专门论题,可以略去.

2) 在一般理论中引入对应于无穷序列 $\{\mathbf{X}_k\}$ 的样本空间. 于是强大数定律可陈述为: $|\mathbf{S}_n - m_n|/n$ 趋向于 0 的概率为 1. 用实变函数的术语来说, 强大数定律所作出的论断是几乎处处收敛,而弱大数定律则等价于按测度收敛.

柯尔莫果洛夫准则. 级数

$$\sum \frac{\sigma_k^2}{k^2} \tag{7.2}$$

收敛是方差为 σ_k^2 的相互独立的随机变量 $\mathbf{X}_k$ 的序列服从强大数定律的一个充分条件.

证明. 设A_ν表示如下事件：至少存在一个满足关系式 $2^{\nu-1}<n\leqslant 2^\nu$ 的 n，使得不等式(7.1)不成立. 显然，只需证明对于所有充分大的 $\nu(\nu>\log N)$ 及一切 r 有

$$\mathbf{P}\{A_\nu\}+\mathbf{P}\{A_{\nu+1}\}+\cdots+\mathbf{P}\{A_{\nu+r}\}<\delta,$$

亦即级数 $\sum\mathbf{P}\{A_\nu\}$ 收敛. 易知，事件 A_ν 蕴含如下事件：对某个满足关系式 $2^{\nu-1}<n\leqslant 2^\nu$ 的 n 有

$$|\mathbf{S}_n-m_n|\geqslant\frac{\varepsilon}{2}\cdot 2^\nu, \tag{7.3}$$

由柯尔莫果洛夫不等式(第九章第 7 节)得

$$\mathbf{P}\{A_\nu\}\leqslant 4\varepsilon^{-2}s_{2^\nu}^2\cdot 2^{-2\nu}. \tag{7.4}$$

故有

$$\sum_{\nu=1}^{\infty}\mathbf{P}\{A_\nu\}\leqslant 4\varepsilon^{-2}\sum_{\nu=1}^{\infty}2^{-2\nu}\sum_{k=1}^{2^\nu}\sigma_k^2$$

$$=4\varepsilon^{-2}\sum_{k=1}^{\infty}\sigma_k^2\sum_{2^\nu\geqslant k}2^{-2\nu}\leqslant 8\varepsilon^{-2}\sum_{k=1}^{\infty}\frac{\sigma_k^2}{k^2}, \tag{7.5}$$

于是定理得证.

作为一个典型的应用我们证明如下的

定理. 如果相互独立的随机变量序列 $\mathbf{X}_k$ 具有共同的分布 $\{f(x_j)\}$ 且 $\mu=\mathbf{E}(\mathbf{X}_k)$ 存在，则序列 $\{\mathbf{X}_k\}$ 服从强大数定律.

这个定理显然比第 1 节中的弱大数定律强. 由于证明在方法论上的重要性，我们对这两个定理独立地进行讨论. 本定理的逆定理可参见问题 17.

证. 我们再次应用截尾法.引进两个新的随机变量序列如下：

$$\mathbf{U}_k=\mathbf{X}_k,\quad \mathbf{V}_k=0,\quad 如果\ |\mathbf{X}_k|<k,$$

$$\mathbf{U}_k=0,\quad \mathbf{V}_k=\mathbf{X}_k,\quad 如果\ |\mathbf{X}_k|\geqslant k. \tag{7.6}$$

$\mathbf{U}_k$ 是相互独立的，且我们将证明它们满足柯尔莫果洛夫准则．显

$$\sigma_k^2 \leqslant \mathbf{E}(\mathbf{U}_k^2) = \sum_{|x_j|<k} x_j^2 f(x_j). \tag{7.7}$$

然，作为简写，令

$$a_\nu = \sum_{\nu-1\leqslant|x_j|<\nu} |x_j| f(x_j). \tag{7.8}$$

因为 $\mathbf{E}(\mathbf{X}_k)$ 存在，故级数 $\sum a_\nu$ 收敛．此外，由 (7.7) 有

$$\sigma_k^2 \leqslant a_1 + 2a_2 + 3a_3 + \cdots + ka_k \tag{7.9}$$

及

$$\sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} \leqslant \sum_{k=1}^{\infty} \frac{1}{k^2} \sum_{\nu=1}^{k} \nu a_\nu = \sum_{\nu=1}^{\infty} \nu a_\nu \sum_{k=\nu}^{\infty} \frac{1}{k^2} \leqslant 2 \sum_{\nu=1}^{\infty} a_\nu < \infty. \tag{7.10}$$

最后，

$$\mathbf{E}(\mathbf{U}_k) = \mu_k = \sum_{|x_j|<k} x_j f(x_j), \tag{7.11}$$

故 $\mu_k \to \mu$ 且 $(\mu_1 + \mu_2 + \cdots + \mu_n)/n \to \mu$．对 $\{\mathbf{U}_k\}$ 应用强大数定律，我们得到如下结论：不等式

$$\left| n^{-1} \sum_{k=1}^{n} \mathbf{U}_k - \mu_k \right| < \varepsilon \tag{7.12}$$

对一切 $n \geqslant N$ 成立的概率不小于 $1-\delta$．现在只需证明，$\mathbf{V}_n$ 可略去不计，即当 $N \to \infty$ 时有一个或一个以上的 $V_n (n > N)$ 异于零的概率趋向于 0．易知，这等价于有无穷多个 n 使得 $\mathbf{V}_n \neq 0$ 的概率为 0[1]，于是由波雷尔-康特立第一引理（第八章第 3 节）知，为此只需证明 $\sum \mathbf{P}\{\mathbf{V}_n \neq 0\}$ 收敛．我们有

$$\begin{aligned} \mathbf{P}\{\mathbf{V}_n \neq 0\} &= \sum_{|x_j|\geqslant n} f(x_j) \\ &\leqslant \frac{a_{n+1}}{n} + \frac{a_{n+2}}{n+1} + \frac{a_{n+3}}{n+2} + \cdots \end{aligned} \tag{7.13}$$

1) 这句话是译者加的 ——译者注．

于是

$$\sum \mathbf{P}\{\mathbf{V}_n \neq 0\} \leqslant \sum_{n=1}^{\infty} \sum_{\nu=n}^{\infty} \frac{a_{\nu+1}}{\nu}$$

$$= \sum_{\nu=1}^{\infty} \frac{a_{\nu+1}}{\nu} \sum_{n=1}^{\nu} 1 = \sum_{\nu} a_{\nu+1} < \infty, \qquad (7.14)$$

证毕。

8. 问　　题

1．证明在例（5.a）中当 $\lambda \leqslant 0$ 时大数定律也成立。当 $\lambda \geqslant -\frac{1}{2}$ 时中心极限定理成立。

2．试判定对于下列相互独立的随机变量的序列 $\mathbf{X}_k$ 大数定律与中心极限定理是否成立，其中 $\mathbf{X}_k$ 的分布定义如下：

（a）$\mathbf{P}\{\mathbf{X}_k = \pm 2^k\} = \frac{1}{2}$；

（b）$\mathbf{P}\{\mathbf{X}_k = \pm 2^k\} = 2^{-(2k+1)}$，$P\{X_k = 0\} = 1 - 2^{-2k}$；

（c）$\mathbf{P}\{\mathbf{X}_k = \pm k\} = \frac{1}{2} k^{-1/2}$，$P\{X_k = 0\} = 1 - k^{-1/2}$。

3．李雅普诺夫条件（1901）。设对于某个固定的 $\delta > 0$ 我们有 $\mathbf{E}(|\mathbf{X}_k|^{2+\delta}) = \lambda_k$。其中 $\lambda_k / s_k^2 <$ 常数。试证明当 $s_n \to \infty$ 时林德堡条件成立。

4．设 $\{X_k\}$ 是相互独立的随机变量序列，其中 $\mathbf{X}_k$ 取每个值

$$0, \ \pm \frac{1}{2k+1} L_k, \ \pm \frac{2}{2k+1} L_k, \ \cdots, \ \pm \frac{k}{2k+1} L_k$$

的概率都是 $1/(2k+1)$。试求加在 $\{L_k\}$ 上的、使 $\{\mathbf{X}_k\}$ 满足弱大数定律及（或）中心极限定理的充分条件。

5．设 $\mathbf{X}_k$ 取值 a_k，$-a_k$ 与 0 的概率分别为 p_k，p_k 与 $1-2p_k$，试讨论同样的问题。

注意：以下七个问题讨论相依变量的弱大数定律。

6．在第五章问题 13 中，如果第 k 次投掷出现红面则令 $\mathbf{X}_k = 1$，否则令 $\mathbf{X}_k = 0$。证明大数定律不成立。

7. 设 $\{\mathbf{X}_k\}$ 相互独立且具有有限的均值 μ 和有限的方差. 试证明对于序列 $\{\mathbf{S}_n\}$ 大数定律不成立,而当 $na_n \to 0$ 时,对于 $a_n\mathbf{S}_n$ 大数定律成立.

8. 设 $\{\mathbf{X}_k\}$ 为一随机变量序列,其中 $\mathbf{X}_k$ 可以依赖于 $\mathbf{X}_{k-1}$ 与 $\mathbf{X}_{k+1}$ 而不依赖于所有其它的 $\mathbf{X}_j$. 试证明,如果 $\mathbf{X}_k$ 具有界的方差,则大数定律成立.

9. 如果对于每个 n, 规定了 $(\mathbf{X}_1, \cdots, \mathbf{X}_n)$ 的联合分布,使得方差有界且所有的协方差为负,则大数定律成立.

10. (续上). 用当 $|j-k| \to \infty$ 时 $\mathrm{Cov}(\mathbf{X}_j, \mathbf{X}_k) \to 0$ 一致地成立的假定来代替 $\mathrm{Cov}(\mathbf{X}_j, \mathbf{X}_k) < 0$ 的条件. 证明大数定律成立.

11. 如果 $|\mathbf{S}_n| < cn$ 且 $\mathrm{Var}(\mathbf{S}_n) > \alpha n^2$, 则对于 $\{\mathbf{X}_k\}$ 大数定律不成立.

12. 在波利亚罐子模型中[第五章(2.c)]如果第 k 次抽得黑球则令 $\mathbf{X}_k$ 等于1, 如果抽得红球则令 $\mathbf{X}_k$ 等于0. 则 $\mathbf{S}_n$ 为 n 次抽取中抽得黑球的总数. 证明对于 $\{\mathbf{X}_k\}$ 大数定律不成立. (提示: 利用问题11及第九章问题30).

---

13. 设 $\{\mathbf{X}_k\}$ 是相互独立的随机变量序列, 其中 $\mathbf{X}_k$ 取值 $r = 2, 3, 4, \cdots$ 的概率为 $p_r = c/(r^2 \log r)$, 此处 c 是使得 $\sum p_r = 1$ 的常数. 试证明如果令 $e_n = c \cdot n \log\log n$, 则广义大数定律成立.

14. 设 $\{\mathbf{X}_n\}$ 是相互独立的随机变量序列, 其中 $\mathbf{X}_n = \pm 1$ 的概率均为 $(1-2^{-n})/2$, $\mathbf{X}_n = \pm 2^n$ 的概率均为 2^{-n-1}. 证明对于 $\{\mathbf{X}_n\}$ 弱大数定律与强大数定律都成立. (注: 这证明条件(5.6)不是必要的).

15. **不利于赌徒的"公平"赌博的例子.** 设在每次试验中赢利的可能值为 $0, 2, 2^2, 2^3, \cdots$; 赢利为 2^k 的概率为

$$p_k = \frac{1}{2^k k(k+1)}, \tag{8.1}$$

而赢利为0的概率为 $p_0 = 1-(p_1+p_2+\cdots)$. 赢利的期望值为

$$\mu = \sum 2^k p_k = \left(1-\frac{1}{2}\right) + \left(\frac{1}{2}-\frac{1}{3}\right) + \left(\frac{1}{3}-\frac{1}{4}\right) + \cdots = 1. \tag{8.2}$$

设对于每次试验赌徒需付的入场费为1, 于是在 n 次试验后他的净赢利(或损失)为 $\mathbf{S}_n - n$. 试证明, 对于每个 $\varepsilon > 0$, 在 n 次试验中赌徒遭受的损失超过 $(1-\varepsilon)n/\log_2 n$ 的概率趋向于1, 此处 $\log_2 n$ 表示以2为底的对数. 即要证明

$$P\left\{S_n - n < -\frac{(1-\varepsilon)n}{\log_2 n}\right\} \to 1. \tag{8.3}$$

提示：利用第4节中的截尾法，但用 $n/\log_2 n$ 来代替（4.2）中的界限 $n \log n$。证明 $U_k = X_k$ 对所有的 $k \leqslant n$ 成立的概率趋向于1，并证明

$$P\left\{|U_1 + \cdots + U_n - nE(U_1)| < \frac{\varepsilon n}{\log_2 n}\right\} \to 1. \tag{8.4}$$

$$1 - \frac{1}{\log_2 n} \geqslant E(U_1) \geqslant 1 - \frac{1+\varepsilon}{\log_2 n}. \tag{8.5}$$

详见[3]。

16. 设 $\{X_n\}$ 是独立同分布的随机变量序列。设 X_n 不具有有限的期望值，并令 A 为正常数。试证明在事件 $|X_n| > An$ 中有无限多个出现的概率为1。

17. **强大数定律之逆**。在问题16的假定下，$|S_n| > A_n$ 对于无穷多个 n 成立的概率为1。

18. **柯尔莫果洛夫准则之逆**。如果 $\sum \sigma_k^2/k^2$ 发散，则存在方差为 $\mathrm{Var}\{X_k\} = \sigma_k^2$ 的相互独立的随机变量的序列 $\{X_k\}$，使得对于这个序列强大数定律不成立。（提示：首先证明级数 $\sum P\{|X_n| > \varepsilon n\}$ 收敛是强大数定律成立的必要条件）。

第十一章　取整数值的变量·母函数

1. 概　　论

在离散随机变量中仅取整数值 $k=0, 1, 2, \cdots$ 的变量特别重要. 它们的研究被母函数的方法所促进，我们将看到，这种方法是特征函数方法的一个特殊情况，后者在概率论中起着重要作用. 更一般地说，母函数的理论属于在微分方程与积分方程中有着广泛应用的运算方法的领域. 自从拉普拉斯与德模佛以来，母函数就已在概率论中得到应用，但这种方法的效能与潜力却很少被充分利用.

定义. 设 $a_0, a_1, a_2, \cdots$ 是实数的一个序列. 如果

$$A(s)=a_0+a_1s+a_2s^2+\cdots \tag{1.1}$$

在某个区间 $-s_0<s<s_0$ 中收敛，则称 $A(s)$ 为序列 $\{a_j\}$ 的母函数.

变数 s 本身并无意义. 如果序列 $\{a_j\}$ 有界，则通过和几何级数相比较可知，(1.1) 至少当 $s<1$ 时收敛.

例. 如果对于所有的 j, $a_j=1$, 则 $A(s)=1/(1-s)$. 序列 $(0, 0, 1, 1, 1, \cdots)$的母函数为 $s^2/(1-s)$. 序列 $a_j=1/j!$ 具有母函数 e^s. 对于固定的 n, 序列 $a_j=\binom{n}{j}$ 具有母函数 $(1+s)^n$. 设 $\mathbf{X}$ 是掷一个均匀的骰子时出现的点数，则 $\mathbf{X}$ 的概率分布具有母函数 $(s+s^2+s^3+s^4+s^5+s^6)/6$.

设 $\mathbf{X}$ 为取值 $0, 1, 2, \cdots$ 的一个随机变量，为方便起见，我们引入记号来表示 $\mathbf{X}$ 的分布及其尾部:

$$\mathbf{P}\{\mathbf{X}=j\}=p_j, \quad \mathbf{P}\{\mathbf{X}>j\}=q_j. \tag{1.2}$$

于是

$$q_k = p_{k+1} + p_{k+2} + \cdots, \quad k \geqslant 0. \tag{1.3}$$

序列 $\{p_j\}$ 与 $\{q_k\}$ 的母函数为

$$P(s) = p_0 + p_1 s + p_2 s^2 + p_3 s^3 + \cdots, \tag{1.4}$$

$$Q(s) = q_0 + q_1 s + q_2 s^2 + q_3 s^3 + \cdots. \tag{1.5}$$

由于 $P(1) = 1$，故 $P(s)$ 的级数至少在区间 $-1 \leqslant s \leqslant 1$ 中绝对收敛. $Q(s)$ 的系数小于 1，故 $Q(s)$ 的级数至少在开区间 $-1 < s < 1$ 中绝对收敛.

定理 1. 当 $-1 < s < 1$ 时我们有

$$Q(s) = \frac{1 - P(s)}{1 - s}. \tag{1.6}$$

证. $(1-s) \cdot Q(s)$ 中 s^n 的系数当 $n \geqslant 1$ 时为 $q_n - q_{n-1} = -p_n$，当 $n = 0$ 时为 $q_0 = p_1 + p_2 + \cdots = 1 - p_0$. 故 $(1-s) \cdot Q(s) = 1 - P(s)$，证毕.

下面我们来研究导数

$$P'(s) = \sum_{k=1}^{\infty} k p_k s^{k-1}. \tag{1.7}$$

这个级数至少当 $-1 < s < 1$ 时收敛. 当 $s = 1$ 时右边形式上化为 $\sum k p_k = \mathbf{E}(\mathbf{X}_k)$. 如果这个期望值存在，则导数 $P'(s)$ 在闭区间 $-1 \leqslant s \leqslant 1$ 上连续. 如果 $\sum k p_k$ 发散，则当 $s \to 1$ 时 $P'(s) \to \infty$. 在这种情况，我们说 $\mathbf{X}$ 具有无限的期望值并记为 $P'(1) = \mathbf{E}(\mathbf{X}) = \infty$. （因为所有的量都是正的，故我们可以放心地使用符号 ∞.）对(1.6)的右边应用中值定理可得 $Q(s) = P'(\sigma)$，其中 σ 是 s 与 1 之间的一个点. 当 $s \to 1$ 时函数 $Q(s)$ 单调地增加，故 $Q(s) \to \mathbf{E}(\mathbf{X})$（有限或无限）. 这就证明了如下的

定理 2. 对于 $\mathbf{E}(\mathbf{X})$ 我们有两种表示法，

$$\mathbf{E}(\mathbf{X}) = \sum_{j=1}^{\infty} j p_j = \sum_{k=0}^{\infty} q_k. \tag{1.8}$$

用母函数来表示，则为

$$\mathbf{E}(\mathbf{X}) = P'(1) = Q(1). \tag{1.9}$$

微分 (1.7) 及关系式 $P'(s) = Q(s) - (1 - s) Q(s)$，根据同

样的方法我们得到

$$\mathbf{E}(\mathbf{X}(\mathbf{X}-1))=\sum k(k-1)p_k=P''(1)=2Q'(1)^{1)}. \tag{1.10}$$

加上 $\mathbf{E}(\mathbf{X})-\mathbf{E}^2(\mathbf{X})$，我们得到 $\mathbf{X}$ 的方差，于是有如下的

定理 3. 我们有

$$\begin{aligned}\mathrm{Var}(\mathbf{X})&=P''(1)+P'(1)-P'^2(1)\\&=2Q'(1)+Q(1)-Q^2(1).\end{aligned} \tag{1.11}$$

如果方差无限，从当 $s\to 1$ 时 $P''(s)\to\infty$.

公式(1.9)与(1.11)经常提供计算 $\mathbf{E}(\mathbf{X})$ 与 $\mathrm{Var}(\mathbf{X})$ 的最简单的方法.

2. 褶　积

设 $\mathbf{X}$ 与 $\mathbf{Y}$ 是两个取非负整数值的独立随机变量，其概率分布分别为 $\mathbf{P}\{\mathbf{X}=j\}=a_j$, $\mathbf{P}\{\mathbf{Y}=j\}=b_j$. 事件 $(\mathbf{X}=j,\mathbf{Y}=k)$ 的概率为 a_jb_k. 和 $\mathbf{S}=\mathbf{X}+\mathbf{Y}$ 是一新随机变量，事件 $\mathbf{S}=r$ 是下列互不相容的事件之并：

$$(\mathbf{X}=0,\ \mathbf{Y}=r),\quad (\mathbf{X}=1,\ \mathbf{Y}=r-1),$$
$$(\mathbf{X}=2,\ \mathbf{Y}=r-2),\ \cdots,\ (\mathbf{X}=r,\ \mathbf{Y}=0).$$

故分布 $c_r=\mathbf{P}(\mathbf{S}=r)$ 由下式给出：

$$c_r=a_0b_r+a_1b_{r-1}+a_2b_{r-2}+\cdots+a_{r-1}b_1+a_rb_0. \tag{2.1}$$

由两个序列 $\{a_k\}$ 与 $\{b_k\}$ 得出一个新序列 $\{c_r\}$ 的运算(2.1)经常出现，为方便起见，我们给它起一个名称并引进一种记号.

定义. 设 $\{a_k\}$ 与 $\{b_k\}$ 是任意两个数列(不必为概率分布). 由(2.1)所定义的新序列 $\{c_r\}$ 称为 $\{a_k\}$ 与 $\{b_k\}$ 的褶积[2]，并记之为

$$\{c_k\}=\{a_k\}*\{b_k\}. \tag{2.2}$$

例. (a) 如果对于所有的 $k\geqslant 0$, $a_k=b_k=1$, 则 $c_k=k+1$.

---

1) 原书误为 $2Q(1)$ ——译者注.

2) 有些作者喜欢用德文 faltung 表示褶积(英文为 Convolution ——译者注). 与之等价的法文名词为 Composition.

如果 $a_k=k$, $b_k=1$, 则 $c_k=1+2+\cdots+k=k(k+1)/2$. 最后,如果 $a_0=a_1=1/2$, 当 $k\geqslant 2$ 时 $a_k=0$, 则 $c_k=(b_k+b_{k-1})/2$, 等等.

序列 $\{a_k\}$ 与 $\{b_k\}$ 的母函数为 $A(s)=\sum a_k s^k$ 与 $B(s)=\sum b_k s^k$. 乘积 $A(s)\,B(s)$ 可由 $A(s)$ 与 $B(s)$ 的幂级数逐项相乘而得. 合并 s 的幂次相同的各项,我们发现在 $A(s)\,B(s)$ 的展开式中 s^r 的系数 c_r 由(2.1)给出. 于是我们有如下的

定理. 设 $\{a_k\}$ 与 $\{b_k\}$ 是两个序列, 其母函数为 $A(s)$ 与 $B(s)$, 褶积为 $\{c_k\}$, 则母函数 $C(s)=\sum c_k s^k$ 是如下的乘积:

$$C(s)=A(s)B(s). \tag{2.3}$$

如果 $\mathbf{X}$ 与 $\mathbf{Y}$ 是取非负整数值的两个相互独立的随机变量, 其母函数[1]为 $A(s)$ 与 $B(s)$, 则它们的和 $\mathbf{X}+\mathbf{Y}$ 的母函数为 $A(s)\,B(s)$.

现在设 $\{a_k\}$, $\{b_k\}$, $\{c_k\}$, $\{d_k\}$, $\cdots$ 是任意的序列. 我们首先作出褶积 $\{a_k\}*\{b_k\}$, 然后作出这个新序列与 $\{c_k\}$ 的褶积,余此类推. $\{a_k\}*\{b_k\}*\{c_k\}*\{d_k\}$ 的母函数为 $A(s)\,B(s)\,C(s)\,D(s)$, 这表明构成褶积的次序是无关紧要的. 例如, $\{a_k\}*\{b_k\}*\{c_k\}=\{c_k\}*\{b_k\}*\{a_k\}$, 等等. 于是褶积是一种可结合和可交换的运算(与随机变量的求和完全一样).

在研究独立随机变量 $\mathbf{X}_n$ 的和时, $\mathbf{X}_n$ 具有相同分布的特殊情况特别重要. 如果 $\{a_j\}$ 是 $\mathbf{X}_n$ 的共同概率分布, 则我们用 $\{a_j\}^{n*}$ 表示 $\mathbf{S}_n=\mathbf{X}_1+\cdots+\mathbf{X}_n$ 的分布. 则

$$\{a_j\}^{2*}=\{a_j\}*\{a_j\},\quad \{a_j\}^{3*}=\{a_j\}^{2*}*\{a_j\},\cdots \tag{2.4}$$

一般地

$$\{a_j\}^{n*}=\{a_j\}^{(n-1)*}*\{a_j\}. \tag{2.5}$$

用话来说, $\{a_j\}^{n*}$ 是其母函数为 $A^n(s)$ 的一个数列. 特别, $\{a_j\}^{1*}$ 与 $\{a_j\}$ 是同一序列, $\{a_j\}^{0*}$ 则定义为其母函数为 $A^0(s)=1$ 的序列,即序列(1, 0, 0, 0, $\cdots$).

---

1) 取非负整数值的随机变量的母函数是指其分布的母函数, 即按(1.4)来确定, 原书未明确给出这一定义——译者注.

例．（b）**二项分布．** 二项分布 $b(k;n,p)=\binom{n}{k}p^kq^{n-k}$ 的母函数为

$$\sum_{k=0}^{n}\binom{n}{k}(ps)^k q^{n-k}=(q+ps)^n. \tag{2.6}$$

以上母函数是 $q+ps$ 的 n 次方这一事实表明，$\{b(k;n,p)\}$ 是 n 个独立随机变量的和的分布，其中每个变量 $\mathbf{X}_j$ 分别以概率 q 与 p 取值 0 与 1，其母函数为 $q+ps$．于是

$$\{b(k;n,p)\}=\{b(k;1,p)\}^{n*}. \tag{2.7}$$

$\mathbf{S}_n=\mathbf{X}_1+\cdots+\mathbf{X}_n$ 的表示法我们曾经使用过［例如在第九章例（3.a）与例（5.a）中］．颠倒以上论证中的次序就可得到二项分布的一种新的推导方法．又由乘法的性质 $(q+ps)^m(q+ps)^n=(q+ps)^{m+n}$ 有

$$\{b(k;m,p)\}*\{b(k;n,p)\}=\{b(k;m+n,p)\}, \tag{2.8}$$

这与第六章（10.4）是同一公式． 通过微分 $(q+ps)^n$ 便可得到 $\mathbf{E}(\mathbf{S}_n)=np$ 与 $\mathrm{Var}(\mathbf{S}_n)=npq$ 的一种简单的证明．

（c）**普阿松分布．** 分布 $p(k;\lambda)=e^{-\lambda}\lambda^k/k!$ 的母函数为

$$\sum_{k=0}^{\infty}e^{-\lambda}\frac{(\lambda s)^k}{k!}=e^{-\lambda+\lambda s}, \tag{2.9}$$

由此得

$$\{p(k;\lambda)\}*\{p(k;\mu)\}=\{p(k;\lambda+\mu)\}, \tag{2.10}$$

这与第六章的（10.5）是同一公式． 根据微分法我们又得到普阿松分布的均值与方差都等于 λ 的结论．

（d）**几何分布与负二项分布．** 设 $\mathbf{X}$ 是具有几何分布

$$\mathbf{P}\{\mathbf{X}=k\}=q^kp, \quad k=0,1,2,\cdots \tag{2.11}$$

的随机变量，其中 p 与 q 是满足条件 $p+q=1$ 的正常数．相应的母函数为

$$p\sum_{k=0}^{\infty}(qs)^k=\frac{p}{1-qs}. \tag{2.12}$$

利用第 1 节中的结果容易求得 $\mathbf{E}(\mathbf{X})=q/p$，$\mathrm{Var}(\mathbf{X})=q/p^2$，这

与第九章的公式（3.c）一致.

在伯努利试验序列中第一次成功恰好在 k 次失败后出现（即在第 $k+1$ 次试验出现）的概率为 $q^k p$，故 $\mathbf{X}$ 可以解释为第一次成功所需的等待时间. 精确地说，这种解释涉及一个无限样本空间，形式定义（2.11）与随机变量的语言的优点在于，我们不需担心原本的样本空间的构造. 以上所述对于第 r 次成功所需的等待时间也同样正确. 设 $\mathbf{X}_k$ 表示第（$k-1$）次成功与第 k 次成功之间失败的数目，则 $\mathbf{S}_r = \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_r$ 为第 r 次成功之前失败的总数（于是 $\mathbf{S}_r + r$ 为直到出现第 r 次成功所需的试验次数）. 伯努利试验的概念要求 $\mathbf{X}_k$ 相互独立且具有相同的分布(2.11)，此处我们就用这个性质来定义 $\mathbf{X}_k$. 于是 $\mathbf{S}_r$ 具有母函数

$$\left(\frac{p}{1-qs}\right)^r. \tag{2.13}$$

根据第二章的二项式展开公式（8.7）可立即得出 s^k 的系数为

$$f(k;r,p) = \binom{-r}{k} p^r(-q)^k, \quad k = 0, 1, 2, \cdots. \tag{2.14}$$

由此立即得到 $\mathbf{P}\{\mathbf{S}_r = k\} = f(k;r,p)$，这与第六章第 8 节所导出的第 r 次成功之前失败次数的概率分布公式是一致的. 我们可将这个结果重述如下：分布 $\{f(k;r,p)\}$ 是几何分布的 r 重褶积，即

$$\{f(k;r,p)\} = \{q^k p\}^{r*}. \tag{2.15}$$

到目前为止，在我们的讨论中 r 一直是正整数. 在第六章第 8 节中我们曾指出，当 $r>0$ 不是整数时 $\{f(k;r,p)\}$ 给出负二项分布的定义. 此时母函数仍然由（2.13）来确定，易知，对于任意的 $r>0$ 负二项分布的均值与方差分别为 rq/p 与 rq/p^2 且

$$\{f(k;r_1,p)\} * \{f(k;r_2,p)\} = \{f(k;r_1+r_2,p)\}. \tag{2.16}$$

3. 在伯努利试验的初过与循环时间中的应用

本节主要是给出理论的具体说明. 以后我们将用不同的方法

导出本节中的结果(见第十三章例(3.b)及问题7; 第十四章第5节及问题11与15—17). 本节结果在 $p=1/2$的特殊情况已在第三章中得到. 然而, 下面的推导给出了母函数方法的一个卓越的例子;此外,将不同的方法进行比较是有教益的.

我们考虑成功概率为 p 的伯努利试验, 如果第 k 次试验出现成功, 则令 $\mathbf{X}_k=1$, 否则令 $\mathbf{X}_k=-1$. 于是 $\mathbf{S}_n=\mathbf{X}_1+\cdots+\mathbf{X}_n$ 为 n 次试验中成功比失败多出现的次数. 用生动的赌博语言来说, $\mathbf{S}_n$ 是在前 n 次试验中某甲所获的净赢利. 为方便起见, 令 $\mathbf{S}_0=0$.

(a) 初过. 设某甲决定在第一次获得正的净赢利(其数量必定是一个单位)时退出赌博. 直接列举所有可能情况便可看出,某甲只可能在第 1, 3, 5, 7, $\cdots$ 次试验退出赌博, 其相应的概率为 p, qp^2, $2q^2p^3$, $5q^3p^4$, $\cdots$ 但一般性的规律不易看出. 这些概率的和 σ 等于某甲的净赢利迟早要变为正的概率. 这个量虽然不能用直接的方法求得, 但我们将证明: 如果 $p\geqslant q$, 则 $\sigma=1$; 如果 $p\leqslant q$, 则 $\sigma=p/q$. 等待净赢利增加到 x 单位相当于等待一个单位的净赢利接连增加 x 次. 故某甲的净赢利迟早要达到 x 单位的概率等于 σ^x. 下面我们来计算 σ 及恰好在第 n 次试验时某甲的净赢利第一次达到 x 单位的概率 $\lambda_n^{(x)}$.

更形式地说,我们要求使得 $\mathbf{S}_1\leqslant 0$, $\mathbf{S}_2\leqslant 0$, $\cdots$, $\mathbf{S}_{n-1}\leqslant 0$, $\mathbf{S}_n=1$ 的概率 λ_n. 更一般地说,如果

$$\mathbf{S}_1<x,\ \mathbf{S}_2<x,\ \cdots,\ \mathbf{S}_{n-1}<x,\ \mathbf{S}_n=x, \tag{3.1}$$

则我们说初过点 $x>0$ 出现在第 n 次试验. 这个事件的概率用 $\lambda_n^{(x)}$ 来表示, 为简便起见, 我们令 $\lambda_n^{(1)}=\lambda_n$. 用赌博的语言来说, (3.1) 表示某甲的净赢利在第 n 次试验时第一次达到 x 单位. 初过这一术语是由于在扩散理论中的应用而提出的.

设初过 $x=1$ 出现在第 r 次试验. 以后的试验所产生的新[1]累积净赢利 $\mathbf{S}_1'=\mathbf{X}_{r+1}$, $\mathbf{S}_2'=\mathbf{X}_{r+1}+\mathbf{X}_{r+2}$, $\cdots$, 是独立于前 r 次

1) 原书无"新"字. 而是分别用 cumulative 与 accumulated 两个词(都是"累积"之意)来表示 $\mathbf{S}_k'$ 与 $\mathbf{S}_n$, 以示区别 ——译者注.

试验的．当且仅当 $\mathbf{S}_1' \leqslant 0$，$\cdots$，$\mathbf{S}_{n-r-1}' \leqslant 0$，$\mathbf{S}_{n-r}' = 1$ 时，初过 $x = 2$ 在时刻 n 出现，且这个事件的概率为 λ_{n-r}．换句话说，初过 $x = 1$ 与 $x = 2$ 分别在第 r 次与第 $n > r$ 次试验出现的概率为 $\lambda_n \lambda_{n-r}$．于是我们有如下的结论：在时刻 n 初过 $x = 2$ 的概率为

$$\lambda_n^{(2)} = \lambda_1\lambda_{n-1} + \lambda_2\lambda_{n-2} + \cdots + \lambda_{n-1}\lambda_1. \tag{3.2}$$

令 $\lambda_0 = 0$[1]，容易看出 $\{\lambda_n^{(2)}\} = \{\lambda_n\} * \{\lambda_n\}$ 是 $\{\lambda_n\}$ 的自褶积．引入母函数

$$\lambda(s) = \sum_{n=1}^{\infty} \lambda_n s^n, \quad \lambda^{(x)}(s) = \sum_{n=1}^{\infty} \lambda_n^{(x)} s^n, \tag{3.3}$$

我们有 $\lambda^{(2)}(s) = \lambda^2(s)$，重复以上论证，由归纳法有

$$\lambda^{(x)}(s) = \lambda^x(s). \tag{3.4}$$

于是我们的任务简化为寻求初过 $x = 1$ 的概率．如果 $\mathbf{X}_1 = 1$，则初过 $x = 1$ 在第一次试验发生．如果 $\mathbf{X}_1 = -1$，则在第一次试验后新累积净赢利必须增加二个单位，于是我们有

$$\lambda_1 = p, \quad \lambda_n = q\lambda_{n-1}^{(2)}, \quad n > 1. \tag{3.5}$$

上式显然等价于 $\lambda(s)$ 的一个二次方程：

$$\lambda(s) = ps + qs\lambda^2(s). \tag{3.6}$$

这方程的两个根中有一个在 $s = 0$ 邻近是无界的，另一个根

$$\lambda(s) = \frac{1 - \{1 - 4pqs^2\}^{\frac{1}{2}}}{2qs} \tag{3.7}$$

是 (3.6) 的唯一有界解．

这样我们就求出了所有初过时间的母函数 (3.4)．根据第二章的二项式公式 (8.7) 可以写出各系数

$$\lambda_{2m-1} = \frac{1}{2q}\binom{1/2}{m}(4pq)^m(-1)^{m-1}, \quad \lambda_{2m} = 0, \tag{3.8}$$

不过我们感兴趣的并不是母函数的明确表达式，更富有启发性的是直接从母函数推出有关结论．

首先注意

---

1) 因为我们已经约定 $\mathbf{S}_0 = 0$ ——译者注．

$$\lambda(1)=\frac{1-|p-q|}{2q}, \tag{3.9}$$

故当 $p \geqslant q$ 时 $\lambda(1)=1$，当 $p<q$ 时 $\lambda(1)=p/q$. 于是我们得到如下结论：$\sum\lambda_k$ 等于 1 或 p/q，就看究竟哪一个较小；如果 q 大于 p（这时赌博对某甲不利），则 $\mathbf{S}_n$ 永远为负的概率等于 $(q-p)/q$.

在 $p=q=1/2$ 的对称情况，$\sum\lambda_k=1$；在接连投掷钱币的过程中，某甲确信他迟早将获得正的赢利. 问题是：这需花多长时间？由 $\lambda'(1)=\infty$ 可以断定：在投掷钱币的试验中初过 $x=1$ 所需的试验次数具有无限的期望值. 如果某甲想要在掷钱币的赌博中获得一个单位的赢利并在达到目的后立即退出赌博，则他应料到，所需试验的次数将是巨大的（因而也就需要有巨额的赌本）. 不用说，初过时间的无限期望值与第三章中详细讨论过的掷钱币的出人意料的起伏特征是紧密相关的.

附注. 现在我们已求得 λ_n 的明确表达式，但还需从 (3.3) 或 (3.4) 来计算初过概率 $\lambda_n^{(x)}$. 标准的分析程序是应用复变数方法. 因此注意如下事实是有趣的：反射原理的简单应用使得我们在第三章第 4 节定理 2 中至少能就 $p=q=1/2$ 的对称情况写出 $\lambda_n^{(x)}$ 的明确表达式.（按照第三章中的记号，我们有 $f_{2n}^{(x)}=\lambda_{2n-x}^{(x)}$）. 从 (3.4) 与 (3.7) 可以看出，对于任意的 p 概率 $\lambda_n^{(x)}$ 等于对称情况的相应概率乘以 $(4pq)^{n/2}(p/q)^{x/2}$. 我们看到，最初等的组合分析往往能代替复杂的解析工具并能解决困难的专门问题；仔细领会这一事实是有教益的.

(b) 循环时间. 如果 $\mathbf{S}_1\neq 0$, $\mathbf{S}_2\neq 0$, $\cdots$, $\mathbf{S}_{n-1}\neq 0$, $\mathbf{S}_n=0$（即，如果成功与失败的累积数第一次出现平衡），则我们说第一次返回零出现在第 n 次试验. 设 f_n 表示这个事件的概率.（显然对于所有的 n，$f_{2n+1}=0$. 用直接列举的方法不难求出前几个 f_{2n}：$f_2=2pq$，$f_4=2p^2q^2$，$f_6=4p^3q^3$，$f_8=10p^4q^4$.）

设 $\lambda_n^{(-1)}$ 为在第 n 次试验初过 $x=-1$ 的概率，换句话说，$\lambda_n^{(-1)}$ 是在 $\lambda_n^{(1)}=\lambda_n$ 中交换 p 与 q 而得的量. 易知，在第 n 次试验第一次[1]返回零等价于第 $n-1$ 次试验初过 1 而第 n 次试验出现失败

1) 原书遗漏了"第一次"一词 —— 译者注.

或第 $n-1$ 次试验初过 -1 而第 n 次试验出现成功[1]，故我们有

$$f_n = q\lambda_{n-1} + p\lambda_{n-1}^{(-1)}. \tag{3.10}$$

乘以 s^n 并相加. 注意到 $\{\lambda_n^{(-1)}\}$ 与 $\{\lambda_n\}$ 的母函数可由彼此交换 p 与 q 而得，于是我们有

$$F(s) = \sum f_n s^n = 1 - (1 - 4pqs^2)^{\frac{1}{2}}{}^{2)}. \tag{3.11}$$

我们得到如下结论：成功与失败的累积数迟早将达到平衡的概率 $\sum f_n$ 为 $1 - |p - q|$.

在 $p = q = 1/2$ 的特殊情况我们有 $\sum f_n = 1$，但概率分布 $\{f_n\}$ 具有无限的期望值. 在第三章第 4 节中，我们曾用完全不同的方法计算出 f_{2n}. 注意以下事实是有教益的：第三章的几个定理不需计算也不需 f_{2n} 的明确表达式而可以直接从母函数 $F(s)$ 得到.（见问题 6—10.）

附注. 在概念上，本节的问题与例 (2.d) 中的等待时间问题是相似的. 在伯努利试验的无穷序列样本空间中我们可以考虑被定义为从初过 $r-1$ 到初过 r 所需的试验数（包括初过 r 那次试验在内）的随机变量 $\mathbf{N}_r$. $\{\mathbf{N}_r\}$ 是相互独立且具有共同母函数 $\lambda(s)$ 的随机变量. 和 $\mathbf{N}^{(x)} = \mathbf{N}_1 + \cdots + \mathbf{N}_x$ 为初过 r 的等待时间，其母函数为 $\lambda^x(s)$. 形式上，我们通过用分布来确定随机变量的方法避免了涉及无穷样本空间[3]. 虽然从分析的观点来看理论是严格而完备的（Self-Contained），但把这个自然的无穷样本空间记在心中对直观和概率解释是有好处的.

4. 部分分式展开

对于给定的母函数 $P(s) = \sum p_k s^k$，其系数可通过微分法由显而易见的公式 $p_k = P^{(k)}(0)/k!$ 求得. 但实际上要求出明确的表达式并不总是可能，而且这样的表达式经常相当复杂，因此有必要考虑近似计算的问题. 获得这种近似的最通常的方法是以部分分式的展开为根据的. 由复变数理论可知，很大一类函数都可以这

1) 原书在叙述中没有考虑第 n 次试验的结果；这是一个疏忽 ——译者注.

2) 原书遗漏了指数 1/2 ——译者注.

3) 确切地说，应是由无穷序列组成的非离散样本空间——译者注.

样展开，但此处我们仅就最简单的有理函数类来进行说明。

设母函数具有形式

$$P(s)=\frac{U(s)}{V(s)}, \tag{4.1}$$

此处 $U(s)$ 与 $V(s)$ 是无共同根的多项式. 为简单起见，假定 $U(s)$ 的次数低于 $V(s)$ 的次数 m. 此处，假设方程 $V(s)=0$ 有 m 个不同的(实的或虚的)根 $s_1, s_2, \cdots, s_m$. 于是

$$V(s)=(s-s_1)(s-s_2)\cdots(s-s_m), \tag{4.2}$$

由代数可知 $P(s)$ 可以分解为部分分式：

$$P(s)=\frac{\rho_1}{s_1-s}+\frac{\rho_2}{s_2-s}+\cdots+\frac{\rho_m}{s_m-s}, \tag{4.3}$$

此处 $\rho_1, \rho_2, \cdots, \rho_m$ 是常数. 为了求 ρ_1，将(4.3)两边同乘 s_1-s；当 $s\to s_1$ 时乘积 $(s_1-s)P(s)$ 趋向于 ρ_1. 另一方面，由(4.1)及(4.2)我们得

$$(s_1-s)P(s)=\frac{-U(s)}{(s-s_2)(s-s_3)\cdots(s-s_m)}. \tag{4.4}$$

当 $s\to s_1$ 时分子趋向于 $-U(s_1)$，而分母则趋向于 $(s_1-s_2)(s_1-s_3)\cdots(s_1-s_m)$，后者就是 $V'(s_1)$. 于是 $\rho_1=-U(s_1)/V'(s_1)$. 同样的论证适应于所有的根，故对于 $k\leqslant m$，有

$$\rho_k=\frac{-U(s_k)}{V'(s_k)}. \tag{4.5}$$

不幸的是，将(4.1)写成(4.3)的形式通常需要进行多方面的数值计算. 然而，一旦获得展开式(4.3)，我们就能容易地导出 $P(s)$ 中 s^n 的系数的精确表达式. 置

$$\frac{1}{s_k-s}=\frac{1}{s_k}\cdot\frac{1}{1-s/s_k}. \tag{4.6}$$

当 $|s|<|s_k|$ 时，我们可将最后的分式展开为几何级数

$$\frac{1}{1-s/s_k}=1+\frac{s}{s_k}+\left(\frac{s}{s_k}\right)^2+\left(\frac{s}{s_k}\right)^3+\cdots. \tag{4.7}$$

以上式代入(4.3)，我们就求得 s^n 的系数 p_n

$$p_n = \frac{\rho_1}{s_1^{n+1}} + \frac{\rho_2}{s_2^{n+1}} + \cdots + \frac{\rho_m}{s_m^{n+1}}. \tag{4.8}$$

于是，为了得到 p_n 我们必须首先求出分母的根 $s_1, \cdots, s_m$，然后由(4.5)确定系数 $\rho_1, \cdots, \rho_m$.

在(4.8)中我们有概率 p_n 的一个精确表达式. 为计算 m 个根通常需要付出大量的劳动，故公式(4.8)主要是具有理论上的价值. 幸运的是(4.8)中有一项几乎经常给出良好的近似. 事实上，设 s_1 是一个根，其绝对值比所有其它根的绝对值都要小. 则(4.8)中的第一个分母最小. 显然，当 n 增大时，其它各项所占的比例递减而第一项则起主要作用. 换言之，如果 s_1 是 $V(s) = 0$ 的一个根，其绝对值比所有其它根的绝对值都要小，则当 $n \to \infty$ 时，

$$p_n \sim \frac{\rho_1}{s_1^{n+1}} \tag{4.9}$$

(符号 $\sim$ 表示两边的比趋向于 1). 通常，甚至对于较小的 n 给出的近似也非常好. (4.9)的主要优点在于它仅需计算代数方程的一个根.

容易除去在推导公式(4.9)时所作的限制. 首先，(4.1)中分子的次数可以超过分母的次数 m. 设 $U(s)$ 的次数为 $m + r(r \geqslant 0)$；用除法就可把 $P(s)$ 化为一个 r 次多项式与一个分式 $U_1(s)/V(s)$ 之和，其中 $U_1(s)$ 是次数低于 m 的多项式. 多项式仅影响分布 $\{p_n\}$ 的前 $r+1$ 项，而如上面所阐明的 $U_1(s)/V(s)$ 能展开为部分分式. 于是(4.9)仍成立. 其次，$V(s)$ 无重根的限制也是没有必要的. 由代数可知，每一个有理函数都可以展开为部分分式. 如果 s_k 是 $V(s)$ 的一个二重根，则部分分式展开式(4.3)将包含一个形如 $a/(s-s_k)^2$ 的付加项，这将使得 p_n 的精确表达式(4.8)中增加形如 $a(n+1)s_k^{-(n+2)}$ 的一项. 然而，只要 s_1 是一个单根，这并不会影响渐近公式(4.9). 我们将这个结果写成如下的定理以备将来参考：

定理. 设 $P(s)$ 是一个有理函数，其分母有一个单根 s_1，这个根的绝对值比所有其它根的绝对值小，则 s^n 的系数由 $p_n \sim \rho_1 s_1^{-(n+1)}$

渐近地给出，其中 ρ_1 由 (4.5) 确定。

在 s_1 为重根的情况也有类例的渐近公式.（见问题 25。）

例. (a) 设 a_n 为 n 次伯努利试验中出现偶数次成功的概率. 这个事件等价于：第一次试验出现失败而随后 $n-1$ 次试验中出现偶数次成功或第一次试验出现成功而随后 $n-1$ 次试验中出现奇数次成功. 故

$$a_n = qa_{n-1} + p(1 - a_{n-1}), \quad a_0 = 1. \tag{4.10}$$

乘以 s^n 并相加，我们便得到母函数 $A(s)$ 所满足的关系式

$$A(s) - 1 = qsA(s) + ps(1-s)^{-1} - psA(s).$$

由此有

$$2A(s) = \{1 - s\}^{-1} + \{1 - (q-p)s\}^{-1},$$

$$2a_n = 1 + (q-p)^n. \tag{4.11}$$

我们注意，最后的公式在各方面都比显而易见的解答 $a_n = b(0; n, p) + b(2; n, p) + \cdots$ 更为可取.

(b) 设 q_n 为在一个理想钱币的 n 次投掷中不出现长为 3 的正面连贯的概率.（注意 $\{q_n\}$ 并不是概率分布；如果 p_n 为第一个长为 3 的正面连贯在第 n 次试验出现的概率，则 $\{p_n\}$ 是一个概率分布，而 q_n 则表示其"尾部"，即 $q_n = p_{n+1} + p_{n+2} + \cdots$.）

容易证明 q_n 满足递推公式

$$q_n = \frac{1}{2} q_{n-1} + \frac{1}{4} q_{n-2} + \frac{1}{8} q_{n-3}. \tag{4.12}$$

事实上，仅当试验以 T, HT 或 HHT 开始时，才有可能在 n 次试验中不出现序列 HHH. 随后的试验不出现连贯 HHH 的概率分别为 q_{n-1}, q_{n-2} 与 q_{n-3}, 故 (4.12) 右边的三项正是使得"无连贯 HHH"出现的三种互不相容的方式的概率.

显然 $q_0 = q_1 = q_2 = 1$, 故 q_n 能依次由 (4.12) 算出. 以 s^n 乘两边并相加便可求得母函数 $Q(s) = \sum q_n s^n$. 我们有

$$Q(s) - 1 - s - s^2$$

$$= \frac{s}{2}\{Q(s) - 1 - s\} + \frac{s^2}{4}\{Q(s) - 1\} + \frac{s^3}{8} Q(s)$$

或

$$Q(s)=\frac{2s^2+4s+8}{8-4s-2s^2-s^3}. \tag{4.13}$$

分母有一个实根 $s_1=1.0873778\cdots$ 及两个复根. 当 $|s|<s_1$ 时我们有 $|4s+2s^2+s^3|<4s_1+2s_1^2+s_1^3=8$,当 $|s|=s_1$ 且 $s\neq s_1$ 时这个不等式仍成立. 故其它两个根的绝对值大于 s_1. 于是由 (4.9) 有

$$q_n\sim\frac{1.236840}{(1.0873778)^{n+1}}, \tag{4.14}$$

其中分子等于 $(2s_1^2+4s_1+8)/(4+4s_1+3s_1^2)$. 甚至对于 n 的较小值这个公式给出的近似也非常好. 例如对于 $q_3=0.875$ 与 $q_4=0.8125$,这个公式所给出的近似值分别为 0.8847 与 0.81360. 随着 n 的增大,百分误差逐渐减少,由此式算出的 $q_{12}=0.41626\cdots$ 精确到五位小数.

5. 二 元 母 函 数

设 $\mathbf{X}$, $\mathbf{Y}$ 是两个取整数值的随机变量,其联合分布为

$$\mathbf{P}\{\mathbf{X}=j,\ \mathbf{Y}=k\}=p_{jk},\qquad j,k=0,1,\cdots. \tag{5.1}$$

我们称依赖于两个变数的函数

$$P(s_1,s_2)=\sum_{j,k}p_{jk}\,s_1^j\,s_2^k \tag{5.2}$$

为二元母函数.

头两节中的考虑也适用于现在的情况而不必作本质的修改,指出下列三个可直接由 (5.2) 得出的明显性质就已足够:

(a) 边缘分布 $\mathbf{P}\{\mathbf{X}=j\}$ 与 $\mathbf{P}\{\mathbf{Y}=k\}$ 的母函数分别为 $A(s)=P(s,1)$ 与 $B(s)=P(1,s)$.

(b) $\mathbf{X}+\mathbf{Y}$ 的母函数为 $P(s,s)$.

(c) 当且仅当 $P(s_1,s_2)=A(s_1)\,B(s_2)$ 对于所有的 s_1 与 s_2 成立时,变量 $\mathbf{X}$ 与 $\mathbf{Y}$ 相互独立.

例. (a) **二维普阿松分布.** 显然

$$P(s_1, s_2) = e^{-a_1-a_2-b+a_1s_1+a_2s_2+bs_1s_2}, \quad a_i > 0, \quad b > 0 \tag{5.3}$$

可展开为幂级数，其系数为正且系数的和为 1. 故 $P(s_1, s_2)$ 表示二维概率分布的母函数. 其边缘分布分别为具有均值 a_1+b 与 a_2+b 的普阿松分布，但和 $\mathbf{X}+\mathbf{Y}$ 具有母函数 $e^{-a_1-a_2-b+(a_1+a_2)s+bs^2}$，故它不是普阿松变量. （它是一个复合普阿松分布；见第十二章第 2 节.）

(b) **多项分布**. 考虑一个 n 次独立试验的序列，每次试验分别以概率 p_0, p_1, p_2 出现结果 E_0, E_1, E_2. 如果 $\mathbf{X}_i$ 是出现 E_i 的次数，则 $(\mathbf{X}_1, \mathbf{X}_2)$ 具有母函数为 $(p_0 + p_1s_1 + p_2s_2)^n$ 的三项分布.

*6. 连续性定理

在第六章中我们曾证明，普阿松分布 $\{e^{-\lambda}\lambda^k/k!\}$ 是当 $n\to\infty$ 时具有概率 p 的二项分布的极限形式，其中 p 依赖于 n 且当 $n\to\infty$ 时 $np\to\lambda$. 即 $b(k; n, p)\to e^{-\lambda}\lambda^k/k!$. $\{b(k; n, p)\}$ 的母函数为 $(q+ps)^n = \{1-\lambda(1-s)/n\}^n$. 通过取对数我们立即可看出这个母函数趋向于 $e^{-\lambda(1-s)}$，后者为普阿松分布的母函数. 我们将证明这种情况可以推广到一般；概率分布收敛于一个极限分布的充要条件是对应的母函数收敛. 不幸的是这个定理的应用范围不广，因为最有趣的离散分布的极限形式是连续分布（例如正态分布是二项分布的极限形式）.

连续性定理. 设对于每个固定的 n，序列 $a_{0,n}, a_{1,n}, a_{2,n}, \cdots$ 是一个概率分布，即

$$a_{k,n}\geqslant 0, \quad \sum_{k=0}^{\infty} a_{k,n} = 1. \tag{6.1}$$

使得对于每个固定的 k 当 $n\to\infty$ 时

$$a_{k,n}\to a_k \tag{6.2}$$

成立的充分必要条件是对于满足条件 $0\leqslant s<1$ 的每个 s 有

---

* 后面并没有用到本节的内容.

$$A_n(s) \to A(s). \tag{6.3}$$

此处

$$A_n(s) = \sum_{k=0}^{\infty} a_{k,n} s^k, \quad A(s) = \sum_{k=0}^{\infty} a_k s^k \tag{6.4}$$

表示对应的母函数.

附注. 如果(6.2)成立,则自动地有$0 \leqslant a_k \leqslant 1$, $\sum a_k \leqslant 1$. 故至少对于 $|s| \leqslant 1$ 母函数存在. 然而,极限序列 $\{a_k\}$ 不一定是概率分布; 例如, 如果分布 $\{a_{k,n}\}$ 的前 n 项为 0, 则极限序列恒为 0. $\{a_k\}$ 为概率分布的充分必要条件是 $\sum a_k = 1$ 或 $A(1) = 1$.

证[1]. 首先假定(6.2)成立. 对于固定的 $s(0 < s < 1)$ 与固定的 ε 我们能选择 r 使得 $s^r/(1-s) < \varepsilon$. 于是

$$|A_n(s) - A(s)| \leqslant \sum_{k=0}^{r} |a_{k,n} - a_k| s^k + 2\varepsilon. \tag{6.5}$$

右边的和仅包含有限项,其中每一项都趋向于零. 故当 n 充分大时 $|A_n(s) - A(s)|$ 可以任意小. 其次,设(6.3)成立. 我们将利用如下的众所周知的事实[2]: 从所给的分布序列中一定能够求得一个收敛的子序列. 如果(6.2)不成立,则一定能够选出收敛于不同极限序列 $\{a_k^*\}$ 与 $\{a_k^{**}\}$ 的两个子序列,于是 $\{A_n(s)\}$ 的对应子序列分别收敛于 $A^*(s) = \sum a_k^* s^k$ 与 $A^{**}(s) = \sum a_k^{**} s^k$. 然而,这与假设(6.3)矛盾. 故由(6.3)可推出(6.2).

例. (a) **负二项分布.** 在例(2.d)中我们曾看到,分布 $\{f(k; r, p)\}$ 的母函数为 $p^r(1-qs)^{-r}$. 现在令 λ 固定, 并令 $p \to 1$, $q \to 0$, 使得 $q = \lambda/r$. 则

$$\left(\frac{p}{1-qs}\right)^r = \left(\frac{1-\lambda/r}{1-\lambda s/r}\right)^r. \tag{6.6}$$

---

1) 这个定理是拉普拉斯–斯提杰变换的连续性定理的一个特殊情况, 此处的证明是仿照一般情况的证法来叙述的. 在文献中,母函数的连续性定理的叙述与证明通常附加了不必要的限制.

2) 这个事实是著名的赫利 (Helly) 定理的一个特殊情况,它容易用康托的"对角线法"来证明, 这种方法可在任一本关于集合论的书中找到. (参见 И. П. 那汤松,实变函数论, p. 247 ——译者注.)

通过取对数可以看出，右边趋向于 $e^{-\lambda+\lambda s}$，这是普阿松分布 $\{e^{-\lambda}\lambda^k/k!\}$ 的母函数. 故如果 $r\to\infty$ 且 $rq\to\lambda$，则

$$f(k;r,p)\to e^{-\lambda}\frac{\lambda^k}{k!}. \tag{6.7}$$

(b) **具有可变概率的伯努利试验.** 考虑 n 次独立试验，其中第 k 次试验出现成功和失败的概率分别为 p_k 与 $q_k=1-p_k$. 成功数 $\mathbf{S}_n$ 能够写成 n 个相互独立的随机变量 $\mathbf{X}_k$ 之和 $\mathbf{S}_n=\mathbf{X}_1+\cdots+\mathbf{X}_n$，其中 $\mathbf{X}_k$ 的分布为 $\mathbf{P}\{\mathbf{X}_k=0\}=q_k$，$\mathbf{P}\{\mathbf{X}_k=1\}=p_k$. $\mathbf{X}_k$ 的母函数为 q_k+p_ks，故 $\mathbf{S}_n$ 的母函数为

$$P(s)=(q_1+p_1s)(q_2+p_2s)\cdots(q_n+p_ns). \tag{6.8}$$

作为以上概型的一个应用，我们假定某城市中每个住户在指定的一天发生火灾具有小概率 p_k. 则和 $p_1+\cdots+p_n$ 为一天中火灾的期望数，n 则为住户数. 在第六章中我们曾看到，如果所有的 p_k 都相等且各个住户随机独立，则火灾数是一个近似于普阿松分布的随机变量. 现在我们证明，在 p_k 并不相等的更为现实的假设下上述结论仍成立. 这个结果使我们进一步相信普阿松分布充分地描述了那些可以看成是很多小概率事件（“成功”）的累积效应的现象. 意外事故和电话呼唤是典型的例子.

现在我们采用变量数目增加的模型，其中 p_k 依赖于 n 使得 $p_1+p_2+\cdots+p_n=\lambda$ 为常数. 于是由 (6.8) 有

$$\log P(s)=\sum_{k=1}^{n}\log\{1-p_k(1-s)\}. \tag{6.9}$$

因为 $p_k\to 0$，故我们能利用等式 $\log(1-x)=-x-\theta x$，其中当 $x\to 0$ 时 $\theta\to 0$. 由此有

$$\log P(s)=-(1-s)\left\{\sum_{k=1}^{n}(p_k+\theta_kp_k)\right\}\to-\lambda(1-s), \tag{6.10}$$

故 $P(s)$ 趋向于普阿松分布的母函数. 因此，$\mathbf{S}_n$ 的极限分布为普阿松分布. 我们得到如下结论：对于大的 n 与中等大小的 $\lambda=p_1+p_2+\cdots+p_n$，$\mathbf{S}_n$ 的分布能用普阿松分布来逼近[参见第九章例 (5.b).]

7. 问　　题

1. 设 $\mathbf{X}$ 为具有母函数 $P(s)$ 的随机变量。求 $\mathbf{X}+1$ 与 $2\mathbf{X}$ 的母函数。

2. (续上)。求以下各数列的母函数:

(a) $\mathbf{P}\{\mathbf{X} \leqslant n\}$,　(b) $\mathbf{P}\{\mathbf{X} < n\}$,　(c) $\mathbf{P}\{\mathbf{X} \geqslant n\}$,

(d) $\mathbf{P}\{\mathbf{X} > n+1\}$,　(e) $\mathbf{P}\{\mathbf{X} = 2n\}$。

3. 设 u_n 是伯努利试验中第一个组合 SF 在第 $n-1$ 次与第 n 次试验中出现的概率。试求母函数、均值与方差.

4. 试讨论在第二章第 12 节中哪个公式表示褶积,在什么地方用到了母函数。

5. 将一个骰子掷任意多次。设 a_n 为使得出现的点数之和为 n 的所有可能方式的数目。试证明 $\{a_n\}$ 的母函数为 $\{1-s-s^2-s^3-s^4-s^5-s^6\}^{-1}$。

注意. 问题 6—10 参照投掷钱币的例子中的通常记号,其中它们包含第三章所求得的某些关系的一种直接推导。令 $u_n = \mathbf{P}\{\mathbf{S}_n = 0\}$, $f_n = \mathbf{P}\{\mathbf{S}_1 \neq 0, \cdots, \mathbf{S}_{n-1} \neq 0, \mathbf{S}_n = 0\}$(初次返回); 按定义, $u_0 = 1$, $f_0 = 0$。我们假定 $\{f_n\}$ 具有母函数 $F(s) = 1 - \{1-s^2\}^{1/2}$ 为已知(由第 3 节),此外不作其它假定。几乎不用计算,也不需关于系数的明确公式。

6. $\{u_n\}$ 的母函数为 $U(s) = \{1-s^2\}^{-1/2}$。

7. 直到时刻 $2n$ 不出现 0 的概率与 $\mathbf{S}_{2n} = 0$ 的概率 u_{2n} 相同.

8. $\mathbf{S}_{2n} = 0$ 且所有的和 $\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_{2n}$ 都 $\geqslant 0$ 的概率等于 $2f_{2n+2}$。

9. 第一次变号紧接着第 $2n$ 次试验后出现的概率等于 $2f_{2n+2}$.

10. 在和 $\mathbf{S}_1, \cdots, \mathbf{S}_n$ 中恰有 k 为 0 的概率具有母函数 $F^k(s)\ U(s) \cdot (1+s)$。

---

11. 在 $p > q$ 的伯努利试验序列中设 a_n 为存在指标 $j > n$ 使得 $\mathbf{S}_j = 0$ 的概率。试证明 a_n 的母函数为 $4pq[p-q+(1-4pqs^2)^{1/2}]^{-1}(1+s)$。

12. 在第九章的等待时间的例 (3.d) 中求 $\mathbf{S}_r$(r 固定)的母函数。证明第九章的均值公式 (3.3) 并计算方差.

13. (续上)。下面是推导同一结果的另一种方法。设 $p_n(r) = \mathbf{P}\{\mathbf{S}_r = n\}$。证明递推公式

$$p_{n+1}(r) = \frac{r-1}{N} p_n(r) + \frac{N-r+1}{N} p_n(r-1). \tag{7.1}$$

直接由(7.1)导出母函数.

14. 对于预先指定的 r 个元素(代替任意 r 个元素)解以上两个问题.

15[1]. 我们称直到第一次失败为止的伯努利试验序列为一轮(turn). 试求 n 轮中成功累积数 $\mathbf{S}_r$ 的概率分布与母函数.

16. (续上)设 $\mathbf{R}$ 为直到第 ν 次成功的接连的轮数(即,第 ν 次成功在第 $\mathbf{R}$ 轮中出现). 证明 $\mathbf{P}\{\mathbf{R}=r\}=p^\nu q^{r-1}\binom{r+\nu-2}{\nu-1}$. 求 $\mathbf{E}(\mathbf{R})$ 与 $\mathrm{Var}(\mathbf{R})$.

17. (续上)考虑分别具有概率 p_1, q_1 与 p_2, q_2 的两个伯努利试验序列. 证明 N 次成功所需的轮数相同的概率可以表示为

$$(p_1p_2)^N\sum_{\nu=1}^{\infty}\binom{N+\nu-2}{\nu-1}^2(q_1q_2)^{\nu-1}$$

$$=(p_1p_2)^N(1-q_1q_2)^{1-2N}\sum_{k=0}^{N-1}\binom{N-1}{k}^2(q_1q_2)^k.$$

18. 设 $\{\mathbf{X}_k\}$ 为相互独立的变量,每一个都以概率 $1/a$ 取值 $0, 1, 2, \cdots, a-1$. 令 $\mathbf{S}_n=\mathbf{X}_1+\cdots+\mathbf{X}_n$. 证明 $\mathbf{S}_n$ 的母函数为

$$P(s)=\left\{\frac{1-s^a}{a(1-s)}\right\}^n,$$

因而

$$\mathbf{P}\{\mathbf{S}_n=j\}=\frac{1}{a^n}\sum_{\nu=0}^{\infty}(-1)^{\nu+j+a\nu}\binom{n}{\nu}\binom{-n}{j-a\nu}.$$

(在和式中仅有有限多项异于 0.)

注意. 当 $a=6$ 时我们就得到在掷 n 个骰子时出现的点数和为 $j+n$ 的概率. 这个问题的解曾由德模佛给出.

19. (续上)概率 $P\{\mathbf{S}_n\leqslant j\}$ 的母函数为 $P(s)/(1-s)$,故

$$\mathbf{P}\{\mathbf{S}_n\leqslant j\}=\frac{1}{a^n}\sum_{\nu}(-1)^\nu\binom{n}{\nu}\binom{j-a\nu}{n}.$$

20. 续上: 极限形式. 如果 $a\to\infty$, $j\to\infty$, 使得 $j/a\to x$, 则

---

1) 问题 15—17 与台球游戏有直接的联系. 成功概率 p 是游戏者技术好坏的一种量度. 游戏者一直玩到失败为止. 故他所累积的成功数就是他的那一"轮"的长度. 游戏一直进行到有一人获得 N 次成功为止. 因此问题 15 给出一个游戏者获得 k 次成功所需的轮数的概率分布, 问题 16 给出平均轮数, 问题 17 给出两个游戏者的比赛出现平局的概率. 进一步的讨论可参见[8].

$$\mathbf{P}\{\mathbf{S}_n \leqslant j\} \to \frac{1}{n!} \sum_{\nu} (-1)^{\nu} \binom{n}{\nu} (x-\nu)^n,$$

其中和是对所有满足 $0 \leqslant \nu < x$ 的 ν 来求的。

注意。这个结果是由拉格朗日 (Lagrange) 得到的。在几何概率理论中右边表示具有区间 $(0, 1)$ 上的均匀分布的 n 个独立随机变量之和的分布。

21. 设 u_n 为 n 次伯努利试验中的成功数能被 3 除尽的概率。试求 u_n 的递推关系并由此求出母函数。

22. 续上：另一种方法。设 v_n 与 w_n 分别表示 $\mathbf{S}_n$ 具有形式 $3\nu+1$ 与 $3\nu+2$ 的概率(故 $u_n+v_n+w_n=1$)。试求三个联立递推关系并由此求出母函数所满足的三个方程。

23. 设 $\mathbf{X}$ 与 $\mathbf{Y}$ 是具有母函数 $U(s)$ 与 $V(s)$ 的独立变量。试证明 $\mathbf{P}\{\mathbf{X}-\mathbf{Y}=j\}$ 是 $U(s)\,V(1/s)$ 中 s^j 的系数，其中 $j=0, \pm 1, \pm 2, \cdots$。

24. **矩母函数**。设 $\mathbf{X}$ 是具有母函数 $P(s)$ 的随机变量，设对于某个 $s_0>1$，$\sum p_n s^n$ 收敛。则所有的矩 $m_r=\mathbf{E}(\mathbf{X}^r)$ 都存在，且序列 $m_r/r!$ 的母函数 $F(s)$ 至少当 $|s|<\log s_0$ 时收敛。且

$$F(s)=\sum_{r=0}^{\infty} \frac{m_r}{r!} s^r = P(e^s).$$

注意。$F(s)$ 通常称为矩母函数，虽然实际上它所产生的是 $m_r/r!$。

25. 设 $A(s)=\sum a_n s^n$ 为有理函数 $U(s)/V(s)$，且 s_1 为 $V(s)$ 的一个根，其绝对值比所有其它根的绝对值都要小。如果 s_1 具有重数 r，试证明

$$a_n \sim \frac{\rho_1}{s_1^{n+r}} \binom{n+r-1}{r-1},$$

此处 $\rho_1=-r!\,U(s_1)/V^{(r)}(s_1)$。

26. **二维负二项分布**。试证明当参数值为正时 $p_0^a\{1-p_1 s_1-p_2 s_2\}^{-a}$ 为 $(\mathbf{X}, \mathbf{Y})$ 的分布的母函数，其中 $\mathbf{X}$、$\mathbf{Y}$ 与 $\mathbf{X}+\mathbf{Y}$ 的边缘分布为负二项分布[1]。

1) 贝茨 (G. E. Bates) 与内曼 (J. Neyman) 在研究意外偏斜 (accident proneness) 时曾使用过这种类型的分布．见[9]．

*第十二章 复合分布·分支过程

1. 随机个变量之和

设 $\{\mathbf{X}_k\}$ 是独立同分布的随机变量序列，其共同分布与母函数分别为 $\mathbf{P}\{\mathbf{X}_k = j\} = f_j$ 与 $f(s) = \sum f_i s^i$. 我们常对形如 $\mathbf{S}_N = \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_N$ 的和感到兴趣，此处项数$\mathbf{N}$是一个独立于$\mathbf{X}_j$的随机变量. 设 $\mathbf{P}\{\mathbf{N} = n\} = g_n$ 为 $\mathbf{N}$ 的分布，$g(s) = \sum g_n s^n$ 为其母函数. 由条件概率的基本公式可以求出 $\mathbf{S_N}$ 的分布 $\{h_j\}$:

$$h_j = \mathbf{P}\{\mathbf{S_N} = j\} = \sum_{n=0}^{\infty} \mathbf{P}\{\mathbf{N} = n\}\mathbf{P}\{\mathbf{X}_1 + \cdots + \mathbf{X}_n = j\}. \tag{1.1}$$

如果 $\mathbf{N}$ 仅取有限多个值，则 $\mathbf{S_N}$ 定义在有限多个 $\mathbf{X}_k$ 的样本空间上. 否则 $\mathbf{S_N}$ 的概率定义将涉及无穷序列 $\{\mathbf{X}_k\}$ 的样本空间，但我们仅考虑 $\mathbf{S_N}$ 的概率分布，并以分布 (1.1) 作为由点 0, 1, 2, … 所组成的样本空间上的变量 $\mathbf{S_N}$ 的定义.

对于固定的 n，$\mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_n$ 的分布由 $\{f_i\}$ 的 n 重自褶积给出，故 (1.1) 可写成如下的简洁形式

$$\{h_j\} = \sum_{n=0}^{\infty} g_n \{f_j\}^{n*}. \tag{1.2}$$

利用母函数可将这个公式加以简化. $\{f_j\}^{n*}$ 的母函数为 $f^n(s)$，于是由 (1.2) 易知，$\mathbf{S_N}$ 的母函数由下式给出:

$$h(s) = \sum_{j=0}^{\infty} h_j s^j = \sum_{n=0}^{\infty} g_n f^n(s). \tag{1.3}$$

上式右端可由在 $g(s)$ 的泰勒展式中用 $f(s)$ 代替 s 而得;故它等于 $g(f(s))$. 这就证明了如下的

---

* 后面没有用到本章的内容.

定理. 和$S_N = X_1 + \cdots + X_N$的母函数是复合函数$g(f(s))$.

以下是两个有趣的特殊情况.

(a) 如果 X_i 是伯努利变量，其分布为 $P\{X_i = 1\} = p$, $P\{X_i = 0\} = q$, 则 $f(s) = q + ps$, 故 $h(s) = g(q + ps)$.

(b) 如果 N 具有均值为 t 的普阿松分布，则

$$h(s) = e^{-t+tf(s)}. \tag{1.4}$$

以此为母函数的分布称为复合分布.

如果 X_i 为伯努利变量，N 具有普阿松分布，则 $h(s) = e^{-tp+tps}$; 和 S_N 具有均值为 p 的普阿松分布.

例. (a) 在第六章例 (7.c) 中我们看到，X 射线引起细胞中染色体的分裂；对于一定的剂量和曝光时间个体细胞中的分裂数 N 具有普阿松分布. 一个分裂的染色体恢复原状的概率为 q，而细胞死亡的概率则为 $p = 1 - q$. 此处 S_N 是观察到的分裂数[10]，它具有均值为 tp 的普阿松分布.

(b) 在陷兽井实验[11] 中 g_n 表示一兽类共有 n 只野兽的概率. 如果每只野兽陷入的概率均为 p，则(假定统计独立)被捕获的一类野兽的只数 S_N 是一随机变量，其母函数为 $g(q + ps)$. 以上的叙述可用多种方式加以改变. 例如，设 g_n 为一只昆虫产 n 个卵的概率，而 p 为每个卵成活的概率. 则 S_N 为成活的卵数. 又如，设 g_n 为一个家庭有 n 个小孩的概率，并设小孩的男女性别比为 $p:q$，则 S_N 表示一个家庭中的男孩数.

(c) 每株植物都有大量的种子，但每粒种子成活的概率很小，故有理由假定个别植物的种子的成活数具有普阿松分布. 如果 g_n 表示亲本植物数的分布，则 $g(e^{-\lambda+\lambda s})$ 为成活种子数的母函数.

2. 复合普阿松分布

首先我们考虑两个典型的例子.

例. (a) 设在长为 t 的一段时间内被闪电击中的次数是均值为 λt 的普阿松变量. 如果 $\{f_n\}$ 为个别电击所造成的损失的概率

分布，则(假定统计独立)在长为 t 的一段时间内遭受的总损失的概率分布为复合普阿松分布

$$\{h_j\} = e^{-\lambda t} \sum \frac{(\lambda t)^n}{n!} \{f_j\}^{n*},$$

其母函数为

$$h(s;t) = e^{-\lambda t+\lambda t f(s)}. \tag{2.1}$$

(b) 在生态学中假定一个小地区中野兽窝的个数具有普阿松分布，其均值与这地区的面积成比例. 如果 $\{f_k\}$ 为一兽窝中野兽只数的概率分布，则 (2.1) 为这地区的野兽总数的母函数.

在第六章中我们看到很多依赖于时间或空间的现象服从普阿松分布，前面的例子则表明复合普阿松分布也经常与这类现象有关.

母函数 (2.1) 具有如下的显著性质:

$$h(s;\ t_1 + t_2) = h(s;\ t_1)\, h(s;\ t_2). \tag{2.2}$$

对此可作如下的直观描述. 对于每个长为 t 的一段时间有一个母函数为 $h(s;\ t)$ 的随机变量与之相应，我们称它为这一段时间的基值 (contribution). 两段不相重迭的时间的基值是相互独立的，这意味着当一段时间 $t = t_1 + t_2$ 分割为两部分时，其基值 $\mathbf{X}(t) = \mathbf{X}(t_1) + \mathbf{X}(t_2)$ 也相应地分解为两个独立随机变量之和.

在下一节我们将证明(在取整数值的随机变量中)，仅有复合普阿松分布具有这个性质. 在阐述定理之前我们先来看两个例子.

例. (c) 负二项分布的母函数为

$$h(t;\ s) = \left(\frac{p}{1 - qs}\right)^t \qquad p + q = 1, \tag{2.3}$$

易证 (2.2) 的确成立. 故负二项分布是一个复合普阿松分布；令

$$\lambda = \log\frac{1}{p},\ f(s) = \frac{1}{\lambda}\log\frac{1}{1 - qs},\ f_n = \frac{\lambda q^n}{n}, \tag{2.4}$$

即可将 (2.3) 写成 (2.1) 的形式. 分布 $\{\lambda q^n/n\}$ 称为对数分布.

(d) **多重普阿松分布.** 设我们按所卷入的车辆数将汽车事故分为单车事故、双车事故等类. 并进一步假定单车、双车等事故分

别具有均值为 $\lambda_1 t$, $\lambda_2 t$, $\cdots$ 的普阿松分布，且各类事故是统计独立的。于是在长为 t 的一段时间内发生事故的车辆总数的母函数为

$$e^{-\lambda_1 t(1-s)}\,e^{-\lambda_2 t(1-s^2)}\,e^{-\lambda_3 t(1-s^3)}\cdots. \tag{2.5}$$

这又是一个复合普阿松分布，其中 $\lambda=\sum\lambda_i$, $f_i=\lambda_i/\lambda$. 反之，每个复合普阿松分布都可以改写为(2.5)的形式，故我们也可将复合普阿松分布解为单个事物、成对事物等的累积效应.

3. 无穷可分分布

我们称概率分布 $\{h_i\}$, $i=0, 1, \cdots$ 为无穷可分的，如果对于每个 n 它可表示为另一概率分布 $\{\phi_i\}$ 的 n 重自褶积，亦即，如果它的母函数 $h(s)$ 的一个 n 次根 $h^{1/n}(s)=\phi(s)$ 产生一个概率分布 $\{\phi_i\}$.

注意，如果 $h(s;t)$ 满足(2.2)，则 $h(s,t)=h^n(s;t/n)$，故对于每个 t, $h(s;t)$ 都是无穷可分的. 上节的论断包含在如下定理中(这个定理是列威关于任意概率分布的一个重要的一般定理的特殊情况).

定理. 如果 $\{h_i\}$ 是无穷可分的，则其母函数能写成(2.1)的形式(比方说 $t=1$).

[注意：$h^t(s)=h(s;t)$ 满足(2.2).]

证. 设对于每个 n, $h^{1/n}(s)$ 为一概率母函数. 这仅当 $h(0)=h_0>0$ 时才有可能. 故 $h(s)$ 必定在某区间 $|s|\leqslant a\leqslant 1$ 内为正且在这个区间内我们有 $0<1-h(s)<1$. 由此可知 $\log h(s)=\log(1-\{1-h(s)\})$ 可展开为泰勒级数

$$\log h(s)=\sum_{i=0}^{\infty}\chi_i s^i \qquad -a<s<a. \tag{3.1}$$

令 $s=0$ 即可看出 $\chi_0<0$. 我们要证明所有其它的 χ_i 都是非负的. 用反证法证之，设命题不成立并设 $r\geqslant 1$ 是使得 $\chi_r<0$ 成立的最小足标. 为书写简便起见，令

$$A(s) = \sum_{\nu=1}^{r-1} \chi_\nu s^\nu, \quad B(s) = \sum_{\nu=r+1}^{\infty} \chi_\nu s^\nu, \quad \frac{1}{n} = \varepsilon. \tag{3.2}$$

于是有

$$h^{1/n}(s) = e^{\varepsilon\chi_0} \cdot e^{\varepsilon A(s)} \cdot e^{\varepsilon\chi_r s^r} \cdot e^{\varepsilon B(s)}. \tag{3.3}$$

由假设有 $h^{1/n}(s) = \sum \phi_k s^k$，其中 $\phi_k \geqslant 0$．考虑式中 s^r 的系数 ϕ_r．幂级数 $B(s)$ 仅包含次数大于 r 的各次幂，故它不会影响到 ϕ_r．因此 ϕ_r 是

$$e^{\varepsilon\chi_0}\left(1 + \varepsilon A(s) + \frac{1}{2}\varepsilon^2 A^2(s) + \cdots\right)(1 + \varepsilon\chi_r s^r) \tag{3.4}$$

中 s^r 的系数．因为 $A(s)$ 是次数 $\leqslant r-1$ 的多项式，故我们有

$$\phi_r = e^{\varepsilon\chi_0}[\varepsilon\chi_r + \varepsilon^2 p(\varepsilon)], \tag{3.5}$$

其中 $p(\varepsilon)$ 是 ε 的一个多项式．如果如反证法所假定的 $\chi_r < 0$，则当 ε 充分小时(3.5)的右边将为负，于是 $\phi_r < 0$，而这是不可能的．这就证明了 $\chi_r \geqslant 0$，$r = 1, 2, \cdots$．此外我们有 $h(1) = 1$，故 $\log h(1) = \sum \chi_\nu = 0$，即 $-\chi_0 = \chi_1 + \chi_2 + \cdots$．于是只要令 $-\chi_0 = \lambda$，$f_i = \chi_i/\lambda$ 就可将 $h(s)$ 写成(2.1)的形式(其中 $t = 1$)．

4. 分支过程的例题

我们将讨论一种随机过程，它可作为很多经验过程的简化模型，也可用来说明母函数的用处．这个过程可以描述如下．

我们考虑一种质点，这种质点能产生新的同类质点．我们称最初的一个单独质点为第零代．每个质点恰好产生 k 个新质点的概率为 $p_k(k = 0, 1, 2, \cdots)$；第 n 代质点的直接后裔构成第 $n+1$ 代．每代各个质点彼此独立．我们所关心的是各代质点的数目．

在用随机变量的术语作严格的讨论之前我们先来考虑几个例子．

(a) **核链锁反应**．这个与原子弹有关的应用是大家所熟知的[1]．质点为中子，它们易与其它质点发生偶然碰撞．设 p 为质点

1) 下面的讨论遵循[12]．在那里取消了空间均匀性的假设．

迟早要碰撞一次因而产生 m 个质点的概率；于是 $q=1-p$ 为质点无后裔的概率，亦即质点始终不活动(换句话说，即被消除或吸收)的概率．在这个概型中后裔的数目只能为 0 和 m，其相应的概率分别为 q 与 p(即 $p_0=q$，$p_m=p$，对于所有其它的 j，$p_j=0$)．在最坏的情况，最初的那个质点始终不活动因而过程未曾开始．在最好的情况，第一代有 m 个质点，第二代有 m^2 个质点，等等．如果 p 接近1，则质点的数目很可能异常迅速地增大．从数学上说，这个数目可能无限增大．从物理上来说，当质点数目很大时，裂变的概率不可能始终为常数，统计独立性也不再成立．不过，对于普通的链锁反应来说，可将"质点数无限增大"的数学描述翻译为"爆炸"．

(b) 姓氏的存亡． 这里(如社会习惯中那样)，仅考虑男性后代；他们起着质点的作用，p_k 为一个新生的男孩将成为恰好 k 个男孩的祖先的概率．我们的概型作了两个人为的简化．事实上，生育的多少会受到世俗倾向的影响，故分布 $\{p_k\}$ 实际上是逐代改变的．此外，共同的遗传和环境必定引起兄弟之间的相似，而这是与我们的统计独立性假设相矛盾的．我们可以将模型加以改进以消除这些缺点，而不会影响其本质特征．我们将导出某家族有 k 个第 n 代子孙的概率，特别是谱系灭绝的概率．姓氏存亡问题似乎是用概率方法所研究的第一个链锁反应．首先讨论这个问题的是格尔顿(1889)；有关这个问题的详尽叙述读者可参看洛特卡的书[13]．洛特卡证明：美国的经验资料可以很完美地用分布 $p_0=0.4825$，$p_k=(0.2126)(0.5893)^{k-1}$ $(k\geqslant 1)$ 来描述，除第一项外这是一个几何分布．

(c) 基因与突变． 有机体中的每一个基因(参见第五章第5节)都有机会在 1，2，3，… 个直接后裔中再现，我们的概型描述这个过程，当然忽略了总体内部以及随时间的变化．这个概型在突变[14] 或基因类型的变化的研究中特别有用．一个自发的突变产生一个单一的新型基因，这个基因起着零代质点的作用．这理论可用来估计突变基因的传播与生存的机会．为确定起见，考虑

(遵循 R. A. 费歇)一个植株，它是某 100 颗种子的父体和另 100颗种子的母体．如果总体的大小始终不变，则这 200 颗种子中平均有两颗将成长为植株．每颗种子得到一个特定基因的概率为 1/2．故一个突变基因恰好在 k 个新植株中表现出来的概率相当于概率为 $p=1/200$ 的 200 次伯努利试验中恰出现 k 次成功的概率，于是有理由假定 $\{p_k\}$ 近似地为均值等于 1 的普阿松分布．如果基因携带一种生物学上的优势，则我们得到均值 $\lambda>1$ 的普阿松分布．

(d) 排队问题[15]．分支过程的理论可以用来分析(在邮局、电话等中)排队的起伏现象． 在柜台空着时来到因而不需等待的一个顾客称为祖先；在祖先的服务时间内来到并参加排队的顾客是他的直接后裔．只要排队还在进行，这个过程就持续下去．在这个例子中我们所感兴趣的是直到排队终止时刻的后裔总数．

5. 分支过程中的绝灭概率

为从数学上描述这个过程，令 $\mathbf{X}_n$ 表示第 n 代的大小．由假设，$\mathbf{X}_0=1$，$\mathbf{X}_1$ 具有已给的概率分布 $\{p_k\}$，其母函数为 $P(s)=\sum p_ks^k$．第二代由第一代的 $\mathbf{X}_1$ 个成员的直接后裔组成；换句话说，我们把 $\mathbf{X}_2$ 看作是 $\mathbf{X}_1$ 个相互独立的变量之和，其中每个变量都具有母函数 $P(s)$．根据第 1 节中的定理，$\mathbf{X}_2$ 的母函数为 $P_2(s)=P(P(s))$．类似地，$\mathbf{X}_3$ 为 $\mathbf{X}_1$ 个变量之和，其中每个变量具有与 $\mathbf{X}_2$ 相同的分布，故 $\mathbf{X}_3$ 的母函数为 $P_3(s)=P(P_2(s))$．由归纳法可知，一般说来，第 $n+1$ 代中的质点数 $\mathbf{X}_{n+1}$ 的母函数 $P_{n+1}(s)$ 由下式递推地定义：

$$P_1(s)=P(s),\qquad P_{n+1}(s)=P(P_n(s)).\tag{5.1}$$

在例 (4.a) 中 $P(s)=q+ps^m$；于是 $P_2(s)=q+p(q+ps^m)^m$，$P_3(s)=q+p\{q+p(q+ps^m)^m\}^m$，等等．对于普阿松分布 $P(s)=e^{-\lambda(1-s)}$，$P_2(s)=e^{-\lambda+\lambda e^{-\lambda+\lambda s}}$等等．这些公式并不十分讨人喜欢，但能由此得出重要的结论．

我们来求过程在第 n 代或在此以前终止的概率 x_n，即 $x_n=$

$\mathbf{P}\{\mathbf{X}_n = 0\} = P(0)$. 如果 $p_0 = 0$，则绝灭为不可能，故我们假定 $0 < p_0 < 1$. 由定义易知，x_n 随 n 增大. 这个事实可用解析方法证明如下. 在区间 $0 < s < 1$ 中函数 $P(s)$ 是单增的[1]，又我们有 $x_1 = P(0) = p_0$. 故 $x_2 = P(x_1) > P(0) = x_1$，又由归纳法有 $x_{n+1} = P(x_n) > P(x_{n-1}) = x_n$. 由此可知序列 x_n 单调增加地趋向于某个数 ζ，显然 ζ 满足方程

$$\zeta = P(\zeta). \tag{5.2}$$

如果 $u > 0$ 是方程 $u = P(u)$ 的任一根，则 $x_1 = P(0) < P(u) = u$ 又由归纳法可得 $x_{n+1} = P(x_n) < P(u) = u$，故有 $\zeta \leqslant u$[2]. 因此，x_n 趋向于 (5.2) 的最小正根.

由于 $y = P(s)$ 的图形是凸的，故这曲线与平分线 $y = s$ 最多交于两点. 它们的确在点 (1,1) 处相交，故方程 (5.2) 最多能有一个根 $0 < \zeta < 1$. 如果这样的一个根存在，则差商 $\{1 - P(\zeta)\}/\{1 - \zeta\}$ 等于 1， 于是由中值定理可知在 ζ 与 1 之间存在一个点 x 使得导数 $P'(x) = 1$. 由此知仅当 $P'(1) > 1$ 时 (5.2) 才有可能存在一个根 $\zeta < 1$. 另一方面，如果 $P'(1) \leqslant 1$，则 $\{1 - P(s)\}/\{1 - s\} < 1$ 对一切 $s < 1$ 成立，而这蕴含 $P(s) > 1$；故 $P(s)$ 的图形位于平分线的上方因而 (5.2) 无根. 这就证明：方程 (5.2) 存在一个正根 $\zeta < 1$ 的充分必要条件是 $P'(1) > 1$，且这个根是唯一的. 已知 $P'(1) = \sum kp_k$ 是每个质点的直接后裔数的期望值，故我们有如下的基本结果：

设 $\mu = \sum kp_k$ 为一个单一质点的直接后裔数的期望值. 如果 $\mu \leqslant 1$，则过程将在第 n 代以前终止(即 $\mathbf{X}_n = 0$)的概率趋向于 1. 如果 $\mu > 1$，则 (5.2) 存在唯一的一个正根 $\zeta < 1$，ζ 为过程在有限多代后终止的概率的极限.

差 $1 - \zeta$ 可称为过程无限延长的概率. 通常 x_n 迅速地收敛于 ζ，故一个有穷尽过程差不多仅能延续很少几代. 所以，ζ 实际

---

1) $P(s)$ 在区间 $(0, +\infty)$ 上也是单调增加的，下面证明 $x_n < u$ 时要用到这一事实——译者注.

2) 原书误为 $\zeta < u$ ——译者注.

上就是迅速绝灭的概率．在例 (4.c) 中我们可称 $1-\zeta$ 为突变基因得以传下去的概率．我果我们不从单一质点而是从 r 个质点开始，则所有 r 个谱系都将逐渐灭绝的概率为 ζ^r，而至少有一谱系延续下去的概率则为 $1-\zeta^r$．即使 ζ 比较大，当初始数 r 很大时 $1-\zeta^r$ 仍接近 1．在例 (4.a) 的核连锁反应中情况就是这样，因此我们可以说：如果 $\mu>1$，则爆炸的概率接近 1，如果 $\mu\leqslant 1$，则过程在有限代以后停止的概率为 1.

我们也能求得第 n 代的期望大小 $\mathbf{E}(\mathbf{X}_n)=P_n'(1)$．因为 $P_n(s)=P(P_{n-1}(s))$，故我们有 $P_n'(1)=P'(P_{n-1}(1))\,P_{n-1}'(1)=P'(1)P_{n-1}'(1)=\mu\mathbf{E}(\mathbf{X}_{n-1})$，于是由归纳法得

$$\mathbf{E}(\mathbf{X}_n)=\mu^n. \tag{5.3}$$

故可以预期后代的大小是按指数律增长的．以上的论证可加以引伸．易见，不仅 $P_n(0)\to\zeta$ 而且对于所有的 $s<1$ 有 $P_n(s)\to\zeta$．这意味着 $s, s^2, s^3, \cdots$ 的系数趋向于 0．在很多代以后无后裔的概率接近 ζ，后裔数超过任一预先指定的界限的概率接近 $1-\zeta$；后裔数目大小适中则几乎是不可能的[1]．

6. 问　　题

1．试证明分布 (1.1) 的均值为 $\mathbf{E}(\mathbf{N})\,\mathbf{E}(\mathbf{X})$，方差为 $\mathbf{E}(\mathbf{N})\operatorname{Var}(\mathbf{X})+\operatorname{Var}(\mathbf{N})\,\mathbf{E}^2(\mathbf{X})$．(a) 用母函数的方法，(b) 直接从定义与条件期望的概念来证明．

2．陷兽井问题 [例 (1.b)]．如果 $\{g_n\}$ 为几何分布，则合成分布也是几何分布．如果 $\{g_n\}$ 为对数分布 [参见公式 (2.4)]，则合成分布为具有附加项的对数分布．

3．在 $\mathbf{N}$ 次伯努利试验中(此处 $\mathbf{N}$ 为具有普阿松分布的随机变量)，成功与失败的次数是统计独立的变量．试将这个结果推广到多项分布：(a) 直接推广，(b) 用多元母函数的方法．[参见第九章例 (1.d)．]

---

1) 关于 $\mathbf{X}_n$ 的状态可参见 [16]．

4. **随机化**. 设 $\mathbf{N}$ 具有均值为 λ 的普阿松分布，将 $\mathbf{N}$ 个球随机地放入 n 个盒中. 不用计算证明：恰有 m 个盒空着的概率为

$$\binom{n}{m} e^{-\lambda m/n} [(1-e)^{-\lambda/n}]^{n-m}.$$

5. (续上)[1]. 试证明当 r (一个确定的数)个球随机地放入 n 个盒中时恰有 m 个盒空着的概率等于上式中 $e^{-\lambda}\lambda^r/r!$ 的系数. (a) 讨论与矩母函数的联系(第十一章问题 24). (b) 利用这结果导出第二章的公式 (11.7).

6. **概率分布的混合**. 设 $\{f_i\}$ 与 $\{g_i\}$ 是两个概率分布，$\alpha>0$, $\beta>0$, $\alpha+\beta=1$. 则 $\{\alpha f_i+\beta g_i\}$ 也是概率分布. 讨论它的意义及其与第五章第 2 节中的罐子模型的联系. 推广到两个以上的分布的情况. 复合普阿松分布能否为这样一个混合?

7. 在分支过程中证明 $\mathrm{Var}(\mathbf{X}_{n+1})=\mu\,\mathrm{Var}(\mathbf{X}_n)+\mu^{2n}\sigma^2$: (a) 利用母函数,(b)利用条件期望. 由此得出 $\mathrm{Var}(\mathbf{X}_n)=\sigma^2$ [2] $(\mu^{2n-2}+\mu^{2n-3}+\cdots+\mu^{n-1})$.

8. (续上). 证明当 $n>m$ 时 $\mathbf{E}(\mathbf{X}_n\mathbf{X}_m)=\mu^{n-m}\mathbf{E}(\mathbf{X}_m^2)$.

9. (续上). 证明 $\mathbf{X}_m$, $\mathbf{X}_n$ 的二元母函数为 $P_m(s_1P_{n-m}(s_2))$. 由此证明上题中的论断.

1) 这个推导各种组合公式的巧妙方法是由 C. 多姆 (C. Domb) [17] 中给出的.
2) 原书误为 λ^2 ——译者注.

第十三章　循环事件·更新方程

1. 直观导引和例子

我们将要研究与重复试验有关的某些能够重复出现的事件型样．粗略地说，以下理论中所考虑的是具有如下性质的型样 $\mathscr{E}$：在 $\mathscr{E}$ 每次出现以后，试验重新从头开始，即 $\mathscr{E}$ 出现后的试验是整个实验的重复．接连两次出现 $\mathscr{E}$ 之间的等待时间是相互独立且具有相同分布的随机变量．

最简单的特殊情况是 $\mathscr{E}$ 代表在伯努利试验序列中"出现一个成功"．第一次成功的等待时间具有几何分布；第一次成功出现后，试验重新从头开始，而且在第 r 次与第 $r+1$ 次成功之间的试验数具有同一几何分布．直到第 r 次成功的等待时间是 r 个独立变量之和[第九章例(3.c)]．反之，设在某个团体中逐个抽查其成员的生日，并令 $\mathscr{E}$ 代表"样本中有两个人有相同的生日"．这里 $\mathscr{E}$ 并不是重复的；因为它一旦出现，就将永远保持下去．虽然抽样可以进行到第二个重复生日出现，但第二阶段的抽样并不是第一阶段的重复．样本越大，出现重复生日的可能性也就越大；故第一次重复生日的等待时间越长，就可能使得第一次出现重复生日和第二次出现重复生日的间隔缩短．这两个接连的等待时间不仅没有相同的分布，而且也是统计相依的．这种等待时间不是循环事件理论的研究对象．

当我们考虑伯努利试验中接连两次成功的出现时，出现一种不同类型的现象．型样 SS 第一次出现是容易定义的，但如果 $\mathscr{E}$ 代表"正好只有两个成功接连出现"，则第三次试验可能取消第二次试验；例如，如果前四次试验的结果为 $SSSF$，则 $\mathscr{E}$ 在第二次试验出现，但整个序列并不包含 $\mathscr{E}$．对我们来说，事件"$\mathscr{E}$ 在第 n 次

试验出现"仅依赖于前 n 次试验的结果而与以后无关这一性质是重要的.

下面我们考虑可以应用循环事件理论的几个典型问题.

例. (a) **伯努利试验中的成功连贯**. 术语"长为 r 的成功连贯"曾用多种方法定义过. 一个由接连三次成功所组成的序列是包含 0 个还是包含 1 个或 2 个长为 2 的连贯，这主要是随我们习惯和方便而定，根据不同的目的，可以采用不同的定义. 然而，如果我们要应用循环事件的理论的话，那么长为 r 的连贯的概念必须这样定义，使得每当一个连贯完成后，过程又重新从头开始. 这就是说，我们要采取下面的定义：在 n 个字母 S 与 F 的序列中所包含的长为 r 的成功连贯数，就是这个序列中所包含的互不重迭的、由恰好 r 个 S 所构成的字母串的个数. 在一列伯努利试验序列中，如果第 n 次试验的结果使得序列的连贯加一个，则我们就说一个长为 r 的连贯在第 n 次试验出现. 因此，在 $SSS|SF|SSS|SSS$ 中有三个长为 3 的成功连贯，它们分别在第 3, 8, 11 次试验中出现；有五个长为 2 的成功连贯，它们分别在第 2, 4, 7, 9, 11 次试验出现. 这个定义把理论大大地简化了，因为此时固定长度的成功连贯变成了循环事件. 计算序列中的连贯数相当于计算其中由至少 r 个 S 所组成的成功串的数目，不过 $2r$ 个接连出现的成功要算两次，$3r$ 个接连出现的成功要算 3 次，等等. （在 7、8 两节中我们还将讨论这个问题.）

(b) **计数器问题**. 用于测量宇宙射线和 α-质点的计数器可用如下的简化模型来描述[1]. 设伯努利试验匀速地进行. 计数器被设计能自动记录成功，但在每次记录以后的恰好 $r-1$ 次试验中它是关闭着的. 于是一个成功被记录的充分必要条件是在它前面的 $r-1$ 次试验中计数器没有记录. 计数器的输出表示一列相关试验的结果；每一个记录有一个后效. 然而，每当计数器处于自由状

---

1) 此处所描述的是所谓第一类计数器的离散类似. 在问题 10 中描述了第二类计数器. 有关讨论见[18].

态(没有关闭)时，情况完全一样，且试验重新从头开始．设 $\mathscr{E}$ 代表“在试验结束时计数器处于自由状态”，则我们得到一个典型的循环型样(参看问题 9 与 10 及第十五章问题 13)．

(c) **返回原点**．在成功概率为 p 的一列伯努利试验中，令 $\mathscr{E}$ 代表“成功和失败的累积次数相等．”和以前一样，我们用具有共同分布 $\mathbf{P}\{\mathbf{X}_k = 1\} = p, \mathbf{P}\{\mathbf{X}_k = -1\} = q$ 的独立随机变量 $\{\mathbf{X}_k\}$ 来描述伯努利试验，并设

$$\mathbf{S}_0 = 0, \quad \mathbf{S}_n = \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_n. \tag{1.1}$$

于是 $\mathbf{S}_n$ 为 n 次试验中正面比反面多出现的次数，易知，当且仅当 $\mathbf{S}_n = 0$ 时，$\mathscr{E}$ 在第 n 次试验出现．我们称 $\mathscr{E}$ 为返回原点．设已给 $\mathbf{S}_n = 0$，则其后的部分和

$$\mathbf{S}'_0 = \mathbf{S}_n, \quad \mathbf{S}'_1 = \mathbf{S}_{n+1}, \quad \mathbf{S}'_2 = \mathbf{S}_{n+2}, \cdots \tag{1.2}$$

与原来的序列 $\{\mathbf{S}_k\}$ 服从同一概率关系，而且当 $\{\mathbf{S}'_k\}$ 返回原点时，$\{\mathbf{S}_k\}$ 也返回原点，反之亦成立．

事件“$\mathscr{E}$ 在第 n 次试验第一次出现”(也叫“第一次返回原点发生在第 n 次试验”)由使得

$$\mathbf{S}_1 \neq 0, \quad \mathbf{S}_2 \neq 0, \quad \cdots, \quad \mathbf{S}_{n-1} \neq 0, \quad \mathbf{S}_n = 0 \tag{1.3}$$

成立的序列 $\{\mathbf{X}_k\}$ 的集合来定义．如果这事件发生，则我们说等待时间 T 等于 n，并记(1.3)的概率为 $f_n = \mathbf{P}\{\mathbf{T} = n\}$．$\{f_n\}$ 的前几项容易用直接枚举所有可能序列的方法来计算；显然当 n 为奇数时 $f_n = 0$，$f_2 = 2pq$，$f_4 = 2p^2q^2$，$f_6 = 4p^3q^3$，$f_8 = 10p^4q^4$，$f_{10} = 28p^5q^5$．同一序列 $\{f_n\}$ 也表示 $\mathscr{E}$ 第 r 次与第 $r+1$ 次出现之间的等待时间的概率分布，我们也称 $\{f_n\}$ 为循环时间的分布．(在第九章第 3 节中曾用母函数的方法求出过这个分布．在第三章中讨论了 $p = q = \frac{1}{2}$ 的特殊情况，而公式则适应于一般情况，因为满足(1.3)的结果数并不依赖于 p．在本章中我们给出一个新的、独立的推导．)

(d) **伯努利试验中的梯级点**．保持上例中的记号，我们定义一个新的型样 $\mathscr{E}$：“$\mathscr{E}$ 在第 n 次试验出现，如果 $\mathbf{S}_n$ 超过前面的所

有和”，亦即如果

$$\mathbf{S}_n > 0, \ \mathbf{S}_n > \mathbf{S}_1, \ \mathbf{S}_n > \mathbf{S}_2, \ \cdots, \ \mathbf{S}_n > \mathbf{S}_{n-1}. \tag{1.4}$$

在这种情况，我们称第 n 次试验（或指标 n）为一个梯级点．在由 $-1, 0, 1, 2, 1, 2, 3, 2, 1, 2, 1, 2, 3, 4, 5$ 给出的部分和序列 $\mathbf{S}_1, \mathbf{S}_2, \cdots$ 中（参见第三章图 3），在第 3，4，7，14，15 次试验中出现梯级点，且接连两次出现之间的等待时间分别为 3，1，3，7，1．第 r 次出现 $\mathscr{E}$ 也就是第一次出现值 r，故梯级点可用初过时刻来描述．

如果 $\mathscr{E}$ 在第 n 次试验出现，则过程按下述意义重新从头开始．假定（1.4）成立，则其后的试验数 $n+m$ 为梯级点的充分必要条件是

$$\begin{aligned}&\mathbf{S}_{n+m} > \mathbf{S}_n, \ \mathbf{S}_{n+m} > \mathbf{S}_{n+1}, \ \mathbf{S}_{n+m} > \mathbf{S}_{n+2}, \ \cdots, \\ &\mathbf{S}_{n+m} > \mathbf{S}_{n+m-1}.\end{aligned} \tag{1.5}$$

令

$$\mathbf{S}_k^* = \mathbf{S}_{n+k} - \mathbf{S}_n = \mathbf{X}_{n+1} + \cdots + \mathbf{X}_{n+k}. \tag{1.6}$$

则 $n+m$ 为序列 $\mathbf{S}_1, \mathbf{S}_2, \cdots$ 的一个梯级点的充分必要条件是 m 为 $\{\mathbf{S}_k^*\}$ 的一个梯级点．显然（1.6）中所定义的运算产生原来的样本空间的一个独立的重复，故 $\mathscr{E}$ 可作为循环事件理论的研究对象．值得注意的是，在这种情况下，序列（1.2）本身和原来的序列在概率上并不相同：在 $\mathscr{E}$ 出现 r 次以后，部分和 $\mathbf{S}_i$ 必定接近于 r 而不是0．然而，就 $\mathscr{E}$ 而论，在 $\mathscr{E}$ 出现后的试验是重新从头开始的．

（梯级点提供了一种工具，用它可将初过时间的研究化归为循环事件，亦即化归为独立随机变量的求和．第九章第 3 节中给出了一个直接的（等价的）方法．梯级点的概念能有益地应用于任意随机变量的序列，例如关于一般的反正弦律．）

(e) 在接连投掷一个均匀的骰子的序列中，令 $\mathscr{E}$ 代表“一点，二点，…，六点出现的次数相等．”这里 $\mathscr{E}$ 显然具有循环的特性，而无需进一步解释．

2. 定　义

我们考虑可能结果为 $E_j(j=1, 2, \cdots)$ 的一列重复试验. 并不要求这些实验是相互独立的(因而可以应用于马尔科夫链). 和往常一样，我们假定在原则上试验可以无限地继续下去，而且概率 $\mathbf{P}\{E_{j_1}, E_{j_2}, \cdots, E_{j_n}\}$ 也能够相容地对所有的有限序列给予定义. 设 $\mathscr{E}$ 是有限序列的一个属性;即我们假定对于每个有限序列 $(E_{j_1}, E_{j_2}, \cdots, E_{j_n})$，可以唯一地确定它是否具有这个属性 $\mathscr{E}$. 我们约定，语句"$\mathscr{E}$ 在(有限或无限)序列 $E_{j_1}, E_{j_2}, \cdots$ 中的第 n 个位置处出现"是子序列"$E_{j_1}, E_{j_2}, \cdots, E_{j_n}$ 具有性质 $\mathscr{E}$"的一种简略的说法. 根据这个约定，$\mathscr{E}$ 在第 n 次试验出现仅仅依赖于前 n 次试验的结果. 这就容易明白,当我们说到一个"循环事件 $\mathscr{E}$"的时候,实际上我们所指的是一类由 $\mathscr{E}$ 的出现这一性质所定义的事件. 所以很清楚，$\mathscr{E}$ 本身与其说是一个事件，倒不如说它是描述某种性质的一个术语. 此处我们用语不够确切，但这种情况通常是被允许的;例如,我们常说"一个二维问题",而大家知道,问题本身是没有维度的.

定义 1. 属性 $\mathscr{E}$ 定义一个循环事件,如果

(a) 为使 $\mathscr{E}$ 在序列 $(E_{j_1}, E_{j_2}, \cdots, E_{j_{n+m}})$ 的第 n 个与第 $n+m$ 个位置处出现，其充分必要条件是 $\mathscr{E}$ 在两个子序列 $(E_{j_1}, E_{j_2}, \cdots, E_{j_n})$ 和 $(E_{j_{n+1}}, E_{j_{n+2}}, \cdots, E_{j_{n+m}})$ 的最后出现.

(b) 在这种情况下,我们有

$$\mathbf{P}\{E_{j_1}, \cdots, E_{j_{n+m}}\} = \mathbf{P}\{E_{j_1}, \cdots, E_{j_n}\}\mathbf{P}\{E_{j_{n+1}}, \cdots, E_{j_{n+m}}\}.$$

$\mathscr{E}$ 在序列 $(E_{j_1}, E_{j_2}, \cdots)$ 中的第 n 个位置处第一次出现之类的说法的意义是很显然的，此处不再赘述. 又显然对于每个循环事件都可以定义如下两个数列:

$$u_n = \mathbf{P}\{\mathscr{E} \text{ 在第 } n \text{ 次试验出现}\}, \qquad (2.1)$$
$$f_n = \mathbf{P}\{\mathscr{E} \text{ 在第 } n \text{ 次试验第一次出现}\}.$$

为方便起见，我们令

$$f_0 = 0, \quad u_0 = 1, \tag{2.2}$$

并引进母函数

$$F(s) = \sum_{k=1}^{\infty} f_k s^k, \quad U(s) = \sum_{k=0}^{\infty} u_k s^k. \tag{2.3}$$

注意，$\{u_k\}$ 并不是概率分布；事实上，在某些典型情况我们有 $\sum u_k = \infty$. 然而，由于事件"$\mathscr{E}$ 在第 n 次试验第一次出现"是互不相容的，故有

$$f = \sum_{n=1}^{\infty} f_n \leqslant 1. \tag{2.4}$$

显然，$1-f$ 可以解释为在无限延长的试验序列中 $\mathscr{E}$ 不出现的概率. 如果 $f=1$，则我们可以引入具有如下分布的随机变量 T:

$$\mathbf{P}\{\mathbf{T}=n\} = f_n. \tag{2.5}$$

当 $f<1$ 时，我们仍然使用 (2.5) 中的记号. 在这种情况，$\mathbf{T}$ 是一个非真正的(或有欠缺的)随机变量，它以概率 $1-f$ 不取任何数值.（当然我们也可以规定此时 T 取值 ∞，显然，对于这种情况并不需要新的规则.）

$\mathscr{E}$ 的等待时间，即一直到第一次出现 $\mathscr{E}$ 的试验次数(包括使得 $\mathscr{E}$ 出现的这次试验在内)，是一个具有分布 (2.5) 的随机变量；不过这个随机变量实际上仅仅在无穷序列 $(E_{j_1}, E_{j_2}, \cdots)$ 的空间中才有定义.

根据循环事件的定义，$\mathscr{E}$ 在第 k 次试验第一次出现且在第 n 次试验第二次出现这一事件的概率等于 $f_k \cdot f_{n-k}$. 所以 $\mathscr{E}$ 在第 n 次试验第二次出现的概率等于

$$f_n^{(2)} = f_1 f_{n-1} + f_2 f_{n-2} + \cdots + f_{n-1} f_1. \tag{2.6}$$

上式右边是 $\{f_n\}$ 的自褶积，故 $\{f_n^{(2)}\}$ 表示两个独立随机变量之和的分布，其中每个变量具有分布 (2.5). 更一般地，如果 $f_n^{(r)}$ 是 $\mathscr{E}$ 在第 n 次试验中第 r 次出现的概率，则我们有

$$f_n^{(r)} = f_1 f_{n-1}^{(r-1)} + f_2 f_{n-2}^{(r-1)} + \cdots + f_{n-1} f_1^{(r-1)}. \tag{2.7}$$

这个简单的事实可以归纳在下面的定理之中.

定理. 设 $f_n^{(r)}$ 表示 $\mathscr{E}$ 在第 n 次试验中第 r 次出现的概率. 则 $\{f_n^{(r)}\}$ 是 r 个独立随机变量 $\mathbf{T}_1, \cdots, \mathbf{T}_r$ 之和

$$\mathbf{T}^{(r)} = \mathbf{T}_1 + \mathbf{T}_2 + \cdots + \mathbf{T}_r \tag{2.8}$$

的概率分布,其中每个变量都具有分布 (2.5). 换句话说, 对于固定的 r, 序列 $\{f_n^{(r)}\}$ 有母函数 $F^r(s)$.

特别是,由此可得

$$\sum_{n=1}^{\infty} f_n^{(r)} = F^r(1) = f^r. \tag{2.9}$$

即 $\mathscr{E}$ 迟早出现 r 次的概率等于 f^r (其实这个结论是可以预料到的). 现在我们引入

定义 2. 如果 $f=1$, 则循环事件 $\mathscr{E}$ 称为常返的[1]; 如果 $f<1$, 则 $\mathscr{E}$ 称为非常返的.

对于非常返的 $\mathscr{E}$, 它出现 r 次以上的概率趋向于 0; 而对于常返的 $\mathscr{E}$, 这个概率恒等于 1. 这一事实也可用如下的话来描述: 常返的 $\mathscr{E}$ 出现无穷多次而非常返的 $\mathscr{E}$ 仅能出现有限次的概率为 1. (这个陈述不仅是一种描述, 而且在无限序列 $E_{j_1}, E_{j_2}, \cdots$ 的样本空间来解释时,这是一个正式的定理.)

我们还需要一个定义. 在伯努利试验中返回原点 [例 (1.c)] 仅能在偶数次试验中出现. 此时 $f_{2n+1} = u_{2n+1} = 0$, 故母函数 $F(s)$ 与 $U(s)$ 与其说是 s 的幂级数, 倒不如说是 s^2 的幂级数. 与此类似,在例 (1.e) 中 $\mathscr{E}$ 仅能在第 6, 12, 18, $\cdots$ 次试验中出现. 在这种情况, 我们说 $\mathscr{E}$ 是周期的. 这样的循环事件有很多烦人的性质;在每个实例中情况是很明显的,但在所有一般性定理中则都要提到周期性这个名义上的特殊情况.

定义 3. 如果存在整数 $\lambda > 1$ 使得循环事件仅能在第 λ, 2λ, 3λ, $\cdots$ 次试验中出现 (即当 n 不能被 λ 整除时 $u_n = 0$), 则称 $\mathscr{E}$ 为周期的. 具有上述性质的 λ 中的最大者称为 $\mathscr{E}$ 的周期.

最后我们注意,在无限序列 $E_{j_1}, E_{j_2}, \cdots$ 的样本空间中 $\mathscr{E}$ 的

1) 在第一版中确定的和不确定的术语都使用过, 但在应用于马尔科夫链时, 现在的专门名词更好一些.

第 $r-1$ 次与第 r 次出现之间的试验数是一个确定的随机变量（可能是一个有欠缺的随机变量），它具有 $\mathbf{T}_r$ 的概率分布．换句话说，我们的变量 $\mathbf{T}_r$ 实际上是代表 $\mathscr{E}$ 接连两次出现之间的等待时间．为了不涉及超出本卷范围的非离散样本空间，我们曾用分析方法给出过 $\mathbf{T}_r$ 的定义，但我们希望，这不至掩盖直观质朴的概率背景．利用循环事件的概念可将一类比较一般的随机变量化归为独立随机变量之和．反之，任一概率分布 $\{f_n\}$, $n=1,2,\cdots$ 可以用来定义一个循环事件．我们用下面的例子来证明这一论断．

例．考虑一个电灯炮、一段保险丝或另外任一种寿命有限的零件．当第一个用坏了时，第二个同种新零件就立刻被换上，第二个用坏了时，第三个又换上去，等等．我们假定另件的寿命是仅取某一单位时间（一年、一天或一秒）的整倍数的随机变量．于是每一时间单位代表一次试验，其可能结果为“更换”和“不更换”．可以把接连的更换看作循环事件．如果 f_n 为一个新零件恰好能用 n 个时间单位的概率，则 $\{f_n\}$ 为循环时间的分布．如果另件的寿命必然是有限的，则 $\sum f_n=1$，此时循环事件是常返的．通常，事先能够肯定零件的寿命不能超过某定值 m，在这种情况下，母函数 $F(s)$ 是一个次数不超过 m 的多项式．在应用中，我们希望求得在时刻 n 发生更换的概率 u_n，这个 u_n 可由方程（3.1）来计算．此处我们得到一类由任意分布 $\{f_n\}$ 所定义的循环事件．$f<1$ 的情况并不排除在外，此时 $1-f$ 可以解释为零件永远不坏的概率．

3. 基本关系

在本节中我们继续采用（2.2）—（2.4）中的记号，并且要研究 $\{f_n\}$ 与 $\{u_n\}$ 之间的关系．根据定义，$\mathscr{E}$ 在第 ν 次试验中第一次出现而且又在其后的第 n 次试验中出现的概率为 $f_\nu u_{n-\nu}$．$\mathscr{E}$ 在第 n 次试验中第一次出现的概率为 $f_n=f_nu_0$．因为这些情形是互不相容的，故我们有

$$u_n=f_1u_{n-1}+f_2u_{n-2}+\cdots+f_nu_0,\quad n\geqslant 1. \tag{3.1}$$

我们看出，上式右边是褶积 $\{f_k\}*\{u_k\}$，其母函数为 $F(s)U(s)$；上式左边是缺 u_0 这一项的数列 $\{u_n\}$，故其母函数为 $U(s)-1$. 于是由 (3.1) 可得 $U(s)-1=F(s)U(s)$，所以我们证明了如下的

定理 1. $\{u_n\}$ 与 $\{f_n\}$ 的母函数之间有如下关系：

$$U(s)=\frac{1}{1-F(s)}. \tag{3.2}$$

注意. 当 $|s|<1$ 时，(3.2) 的右边可以展开为收敛的几何级数 $\sum F^r(s)$. $F^r(s)$ 中 S^n 的系数 $f_n^{(r)}$ 为 $\mathscr{E}$ 在 n 次试验中第 r 次出现的概率，而方程 (3.2) 则等价于

$$u_n=f_n^{(1)}+f_n^{(2)}+\cdots. \tag{3.3}$$

这个式子表达了如下的显然事实：如果 $\mathscr{E}$ 在第 n 次试验出现，则它在前面已出现了 0, 1, 2, …, $n-1$ 次. (显然当 $r>n$ 时 $f_n^{(r)}=0$.)

定理 2. $\mathscr{E}$ 为非常返的充要条件是

$$u=\sum_{j=0}^{\infty}u_j \tag{3.4}$$

为有限. 当这条件满足时，$\mathscr{E}$ 迟早总会出现的概率 f 为

$$f=\frac{u-1}{u}. \tag{3.5}$$

注意. 我们可把 u_j 解释为一个随机变量的期望值，当 $\mathscr{E}$ 在第 j 次试验出现时这个随机变量的值为 1，否则为 0. 于是 $u_1+u_2+\cdots+u_n$ 是在 n 次试验中 $\mathscr{E}$ 出现的期望次数，故 $u-1$ 可以解释为在无穷多次试验中 $\mathscr{E}$ 出现的期望次数.

证. 由于系数 u_k 是非负的，故当 $s\to 1$ 时 $U(s)$ 单调增加，因此对于每个 N，

$$\sum_{n=0}^{N}u_n\leqslant\lim_{s\to 1}U(s)\leqslant\sum_{n=0}^{\infty}u_n=u.$$

因为当 $f<1$ 时 $U(s)\to(1-f)^{-1}$，而当 $f=1$ 时 $U(s)\to\infty$，从而就能推出定理.

下面的定理是特别重要的[1]．虽然证明是初等的，但它对于理解问题的概率意义并无帮助，故我们把它放在本章的最后．（不过，可参看问题 1．）

定理 3. 设 $\mathscr{E}$ 是常返的且不是周期的，用 μ 表示循环时间 T_ν 的均值，即

$$\mu = \sum j f_j = F'(1) \tag{3.6}$$

（可能 $\mu = \infty$）．则当 $n \to \infty$ 时，

$$u_n \to \mu^{-1} \tag{3.7}$$

（如果平均循环时间为无穷，则 $u_n \to 0$）．

定理 4. 如果 $\mathscr{E}$ 是常返的且具有周期 $\lambda > 0$，则当 $n \to \infty$ 时

$$u_{n\lambda} \to \lambda\mu^{-1}, \tag{3.8}$$

而当 k 不能被 λ 整除时 $u_k = 0$.

证. 因为 $\mathscr{E}$ 具有周期 λ，故级数 $F(s) = \sum f_n s^n$ 仅包含 s^λ 的各次幂，因而 $F(s^{1/\lambda}) = F_1(s)$，其中 $F_1(s)$ 是具有正系数的幂级数且 $F_1(1) = 1$. 由定理 3 可知，$U_1(s) = \{1 - F_1(s)\}^{-1}$ 的系数趋向于 μ_1^{-1}，此处

$$\mu_1 = F_1'(1) = \lambda^{-1}F'(1) = \lambda^{-1}\mu.$$

（显然 μ 与 μ_1 或者同时为有限或者同时为无限．）现在我们有 $U(s) = U_1(s^\lambda)$，故 (3.8) 成立.

例. (a) 作为一个平凡的例子，令 $\mathscr{E}$ 表示伯努利试验中的"成功"．于是由定义可知 $u_n = p$. 定理 3 所陈述的是，相继两次成功之间的试验期望数为 p^{-1}. 这里 $U(s) = 1 + ps(1 - s)^{-1} = (1 - qs)(1 - s)^{-1}$，于是由定理 1 可知 $F(s) = ps(1 - qs)^{-1}$，这就证明了相继两次成功之间的等待时间具有几何分布.

---

1) 见[19]．这个定理是为了更好的理解柯尔莫果洛夫建立的有限马尔科夫链的遍历性而被猜测到并被证明的．在本书第一版出版后，钟开莱注意到，定理 3 实际上与柯尔莫果洛夫的遍历性定理等价，因而可由它导出．以前有很多论文研究各种特殊情况及其变种．后来，伯拉克威尔 (D. Blackwell)、钟开莱、耳多斯和华尔夫维茨等人用不同的方法将定理 3 推广到连续型随机变量的情况并使之进一步精确化．伯拉克威尔简洁地证明：(3.7) 对均值为正的所有取整数值的随机变量成立(不必像本书那样，限于取正整数值的随机变量)．他的方法是以任意变量的梯级点的应用为基础的[参见例 (1.d)]．见[20]．

(b) **伯努利试验中返回原点** [例(1.c)]． 如果在第 k 次试验时，成功和失败的累积数相等，则 k 必须为偶数；如令 $k=2n$，则在这 k 次试验中，有 n 次试验的结果为成功而其它 n 次试验的结果为失败．于是 u_{2n} 就是 $2n$ 次试验中成功与失败出现的次数相等的概率，故我们有

$$u_{2n}=\binom{2n}{n}p^n q^n. \tag{3.9}$$

由二项分布的正态逼近可知(也容易用史特令公式来证明)：

$$\binom{2n}{n}2^{-2n}\sim\frac{1}{(\pi n)^{1/2}}. \tag{3.10}$$

故有

$$u_{2n}\sim\frac{(4pq)^n}{(\pi n)^{1/2}}, \tag{3.11}$$

此处符号 $\sim$ 表示两边的比趋向于 1．

如果 $p\neq 1/2$，则 $4pq<1$，此时级数 $\sum u_{2n}$ 比公比为 $4pq$ 的几何级数收敛得更快．如果 $p=1/2$，则 $u_{2n}\sim(\pi n)^{-\frac{1}{2}}$；故 $\sum u_{2n}$ 发散，但 $u_{2n}\to 0$．由上面的定理可以推出，以概率 1 成立着：如果 $p\neq q$，则累积和 $\mathbf{S}_n$ 只可能有有限多次等于 0．如果 $p=q=1/2$，则它们将经过 0 无限多次，但平均循环时间是无限的．

在 $p\neq q$ 的情况，上述论断在直观上是很显然的，并且也可由强大数定律推出．用赌博的语言来说，如果赌博对某甲有利，则他可以确信，在最初几次起伏以后他的净赢利将永远为正．当 $p=q=1/2$ 时，上述论断在直观上则很不明显，而这正是在第三章第 7 节中所讨论的掷钱币的起伏问题的佯谬特征的根源．

由上面的定理还可导出几个附加的结论．利用容易验证的公式

$$\binom{2n}{n}=\binom{-\frac{1}{2}}{n}\cdot(-4)^n \tag{3.12}$$

及第二章的二项式展开式 (8.7)，由 (3.9) 可得

$$U(s)=\sum_{n=0}^{\infty}u_{2n}s^{2n}=(1-4pqs^2)^{-\frac{1}{2}}. \qquad (3.13)$$

如果 $p\neq 1/2$，则 $u=U(1)=(1-4pq)^{-\frac{1}{2}}=|p-q|^{-1}$. 由 (3.5) 我们得到如下结论：成功和失败的累积数迟早将要相等的概率为

$$f=1-|p-q|. \qquad (3.14)$$

(这是至少返回原点一次的概率.)

由 (3.2) 可得循环时间的母函数

$$F(s)=1-(1-4pqs^2)^{\frac{1}{2}}. \qquad (3.15)$$

当 $p=q=1/2$ 时,

$$F(s)=1-(1-s^2)^{\frac{1}{2}}, \qquad (3.16)$$

由二项展开式可得

$$f_{2n}=(-1)^{n+1}\binom{\frac{1}{2}}{n}=\frac{1}{n}\binom{2n-2}{n-1}2^{-2n+1} \qquad (3.17)$$

(当 n 为奇数时 $f_n=0$). 方程 (3.17) 给出了投掷钱币的古典赌博中返回原点的循环时间的分布. (在第三章第 4 节中我们曾用不同的方法导出过这个公式. 现在的方法虽然不是最初等的，但却是最简单的.)

(c) **多个钱币的投掷中的相同成功数.** 我们考虑独立重复地投掷两个钱币这一试验，每当两个钱币出现正面的累积次数相等(因而出现反面的累积次数也相等)时，我们说 $\mathscr{E}$ 出现. 显然

$$u_n=\frac{1}{2^{2n}}\left\{\binom{n}{0}^2+\binom{n}{1}^2+\binom{n}{2}^2+\cdots+\binom{n}{n}^2\right\}. \qquad (3.18)$$

利用第二章的 (12.11) 及本章的 (3.10)，我们有

$$u_n=\binom{2n}{n}2^{-2n}\sim\frac{1}{(n\pi)^{\frac{1}{2}}}. \qquad (3.19)$$

故 $\sum u_n$ 发散，但 $u_n\to 0$. 所以 $\mathscr{E}$ 是常返的，但具有无限的平均循环时间.

更一般地，考虑同时投掷 r 个钱币这一试验，并令 $\mathscr{E}$ 表示

所有钱币出现正面的累积数相同这一循环事件．则

$$u_n = \frac{1}{2^{rn}}\left\{\binom{n}{0}^r + \binom{n}{1}^r + \cdots + \binom{n}{n}^r\right\}. \tag{3.20}$$

为了估计 u_n，我们注意，二项分布的最大项小于 $n^{-\frac{1}{2}}$．所以

$$u_n < n^{-\frac{1}{2}(r-1)} 2^{-n}\left\{\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n}\right\} = n^{-\frac{1}{2}(r-1)}. \tag{3.21}$$

故当 $r \geqslant 4$ 时 $\sum u_n$ 收敛．上面我们已看到，当 $r = 2$ 时 $\sum u_n$ 是发散的．$r = 3$ 的情况必须另外考虑．由二项分布的正态逼近可知，当 n 充分大而 k 在 $\frac{1}{2}n - n^{\frac{1}{2}}$ 与 $\frac{1}{2}n + n^{\frac{1}{2}}$ 之间时，我们有 $\binom{n}{k}2^{-n} > cn^{-\frac{1}{2}}$，此处 c 为正常数(比方说 e^{-4})．故当 $r = 3$ 时，

$$u_n > 2n^{\frac{1}{2}}(c^3 n^{-\frac{3}{2}}) = 2c^3/n, \tag{3.22}$$

因而 $\sum u_n$ 发散．换句话说，当且仅当 $k \leqslant 3$ 时，k 个钱币出现正面的累积数相等这一循环事件 $\mathscr{E}$ 是常返的．在每种情况，平均循环时间都是无限的．

(d) **骰子**．在例 (1.e) 中我们考虑过掷骰子时出现一点、两点、三点、…、六点的累积数相等这个循环事件．显然 $\mathscr{E}$ 的周期为 6，且 $u_{6n} = (6n)!\,(n!)^{-6}\,6^{-6n}$．利用史特令公式容易证明，$u_{6n}$ 与 $n^{-\frac{5}{2}}$ 同阶，故 $\sum u_n$ 收敛．因此 $\mathscr{E}$ 是非常返的．由(3.7)容易算出，一次循环的时间约为 0.022．

(e) 在连贯理论中的应用见第 7, 8 两节．

4. 更 新 方 程

循环事件理论中的基本方程 (3.1) 是所谓更新方程 (4.1) 的一个特殊情况，有很多问题与这个更一般的方程有关．下面我们来证明，上节中的定理，不必作很大的更动，就可适用于现在的更一般的情况．本节的讨论是纯分析性的，概率解释和应用将放在下一节．

设 $\{a_n\}$ 与 $\{b_n\}$ 是满足条件 $0 \leqslant a_n < 1$ 及 $b_n > 0$ $(n = 0, 1, 2, \cdots)$ 的两个序列. 第三个序列 $\{u_n\}$ 由如下的递推关系来定义:

$$u_n = b_n + (a_0 u_n + a_1 u_{n-1} + \cdots + a_n u_0) \tag{4.1}$$

或

$$\{u_n\} = \{b_n\} + \{a_n\} * \{u_n\}. \tag{4.2}$$

解 (4.1) 依次可得

$$u_0 = b_0/(1 - a_0), \ u_1 = (b_1 + a_1 u_0)/(1 - a_0), \ \cdots,$$

故 $\{u_n\}$ 的存在唯一性是不成问题的. 我们所关心的是当 $n \to \infty$ 时 $\{u_n\}$ 的状态,对于这个问题有很多著作(大多是争论性的)进行过讨论.

如果令 $b_n = 0$, $a_n = f_n (n = 1, 2, \cdots)$ 以及 $b_0 = 1$, $a_0 = 0$, 则方程 (4.1) 就成为 (3.1), 故更新方程 (4.1) 是更为一般的, 但它的性质可由方程 (3.1) 的性质导出. 我们再一次引入母函数

$$A(s) = \sum a_n s^n, \ B(s) = \sum b_n s^n, \ U(s) = \sum u_n s^n. \tag{4.3}$$

由于系数 a_n 与 b_n 有界, 故前面两个级数至少当 $|s| < 1$ 时收敛; 很快就可看到,第三个级数的收敛性是显然的. 方程 (4.1) 现在可以写成 $U(s) = B(s) + A(s) U(s)$ 的形式,或

$$U(s) = \frac{B(s)}{1 - A(s)}. \tag{4.4}$$

当 $B(s) \equiv 1$ 时, (4.4) 就化归为 (3.2) 的形式,但两者之间还有差别,即现在的 $\{a_n\}$ 未必是循环时间的分布,故 $A(s)$ 不但可以小于 1 而且也可以大于 1.

如果存在正整数 $\lambda > 1$, 使得除 $a_\lambda, a_{2\lambda}, a_{3\lambda}, \cdots$ 可能不为零以外, 所有的 a_k 都等于零, 则我们称 $\{a_n\}$ 为周期的. 此时 $A(s)$ 是 s^λ 的幂级数. 具有这种性质的最大整数 λ 称为周期.

定理 1. 设 $\{a_n\}$ 不是周期的,而且 $B(1) = \sum b_n$ 有限.

(a) 如果 $\sum a_n = 1$, 则

$$u_n \to B(1)\mu^{-1}, \text{ 此处 } \mu = \sum n a_n. \tag{4.5}$$

(特别,如果 $\sum n a_n$ 发散,则 $u_n \to 0$.)

(b) 如果 $\sum a_n < 1$, 则级数

$$\sum u_n = B(1)\{1 - A(1)\}^{-1} \tag{4.6}$$

收敛.

(c) 如果 $\sum a_n > 1$（包括级数发散的情况在内），则方程 $A(x) = 1$ 存在唯一的正根 $x < 1$，且此时有

$$u_n \sim \frac{B(x)}{A'(x)} x^{-n}. \tag{4.7}$$

其中符号 $\sim$ 表示两边的比趋向于 1.（由关系式 (4.7) 可知，u_n 按几何数列增加；因为当 $|s| < 1$ 时 $A(s)$ 是正则的，故导数 $A'(x)$ 为有限.）

证. (a) 设 v_n 是 $\{1 - A(s)\}^{-1}$ 中 s^n 的系数，则由上节定理 3 可知 $v_n \to \mu^{-1}$. 现在

$$u_n = v_n b_0 + v_{n-1} b_1 + \cdots + v_0 b_n. \tag{4.8}$$

对于每个固定的 k，当 $n \to \infty$ 时，项 $v_{n-k} b_k$ 趋向于 b_k/μ. 此外，v_n 是有界的. 由此可以推出，当 N 充分大时，u_n 与

$$u_n' = v_n b_0 + v_{n-1} b_1 + \cdots + v_{n-N} b_N \tag{4.9}$$

之差可以任意小，且 $u_n' \to (b_0 + \cdots + b_N)/\mu$，而这个极限值与 $B(1)/\mu$ 之差也可任意小，于是定理的 (a) 证毕.

(b) 第三节定理 2 的证明无需改变，仍然适用于此处的一般情况.

(c) 这里只要将 (a) 中的结果应用于序列 $\{a_n x^n\}$，$\{b_n x^n\}$ 与 $\{u_n x^n\}$ 即可，它们的母函数分别为 $A(xs)$，$B(xs)$ 与 $U(xs)$，且它们之间的关系与原序列之间的关系 (4.2) 是一样的.

为完备起见，下面我们考虑 $\{a_n\}$ 为周期序列的情况，这时 $A(s) = \sum a_{n\lambda} s^{n\lambda}$ 为 s^λ 的幂级数. 在这种情况下，我们按足标将系数 u_n 分成几个同余类(模为 λ)：$\{u_0, u_\lambda, u_{2\lambda}, u_{3\lambda}, \cdots\}$，$\{u_1, u_{\lambda+1}, u_{2\lambda+1}, u_{3\lambda+1}, \cdots\}$，$\cdots$，$\{u_{\lambda-1}, u_{2\lambda-1}, u_{3\lambda-1}, \cdots\}$.

由 (4.4)[1] 可知，系数 $u_{n\lambda}$ 仅依赖于 $b_0, b_\lambda, b_{2\lambda}, \cdots$，而不依赖于其足标不能被 λ 整除的那些 b_k. 因此我们将 $U(s)$ 与 $A(s)$ 表

1) 直接由 (4.1) 来看更明显一些 ——译者注.

成 λ 个 s^λ 的幂级数之和:

$$\begin{aligned} U(s) &= U_0(s) + sU_1(s) + \cdots + s^{\lambda-1}U_{\lambda-1}(s), \\ B(s) &= B_0(s) + sB_1(s) + \cdots + s^{\lambda-1}B_{\lambda-1}(s), \end{aligned} \tag{4.10}$$

其中

$$U_j(s) = \sum_{n=0}^{\infty} u_{n\lambda+j}\, s^n, \quad B_j(s) = \sum_{n=0}^{\infty} b_{n\lambda+j}\, s^n. \tag{4.11}$$

于是由 (4.4), 对于一切 $j = 0, 1, \cdots, \lambda - 1$, 有

$$U_j(s) = \frac{B_j(s)}{1 - A(s)}. \tag{4.12}$$

这里所有的函数都是 s^λ 的幂级数, 因此在作变换 $s^\lambda = t$ 后就可应用前面的定理. 由此得到

定理 2. 在周期为 λ 的情况下, 序列 $\{u_n\}$ 是渐近周期的; 如果 $A(1) = 1$, 则 λ 个子序列 $\{u_{n\lambda+j}\}$ 都有一个极限

$$\lim_{n\to\infty} u_{n\lambda+j} = \frac{\lambda B_j(1)}{\mu}, \tag{4.13}$$

其中

$$B_j(1) = b_j + b_{\lambda+j} + b_{2\lambda+j} + b_{3\lambda+j} + \cdots.$$

例. 累次平均. 设 u_1, u_2, u_3 是三个已给正数, 用接连取算术平均值的方法构造一个无穷数列 $\{u_n\}$:

$$u_4 = \frac{1}{3}(u_1 + u_2 + u_3), \quad u_5 = \frac{1}{3}(u_2 + u_3 + u_4), \cdots$$

$$u_{n+3} = \frac{1}{3}(u_n + u_{n+1} + u_{n+2}), \cdots. \tag{4.14}$$

我们来求 $\{u_n\}$ 的渐近性质. 更精确地说, 我们将证明

$$u_n \to \frac{1}{6}(u_1 + 2u_2 + 3u_3). \tag{4.15}$$

不用说, 此处的方法可以应用到任意的平均值上去(参见本章问题 5 及第十五章问题 15). 关键之处是, 这类问题可化归为更新方程 (4.1), 因而可从一个新的角度来阐明其性质.

如果我们令

$$a_0 = 0,\ a_1 = a_2 = a_3 = \frac{1}{3},\ a_n = b_n = 0 \quad (n \geqslant 4), \quad (4.16)$$

则当 $n \geqslant 4$ 时,(4.14)与(4.1)是一致的．为了对一切 n 把(4.14)化归为(4.1)，我们必须定义 $b_0 = u_0 = 0$，并由下式来确定 b_1, b_2, b_3:

$$b_1 = u_1,\ b_2 = u_2 - \frac{1}{3}u_1,\ b_3 = u_3 - \frac{1}{3}(u_1 + u_2). \quad (4.17)$$

现在由定理 1(a) 即可得到(4.15)而无需进一步的计算．因为母函数 $U(s)$ 是有理函数，故为了看出(4.15)中的极限按指数的速度趋近并估计两边的差,我们可将它展开为部分分式．

5. 延迟了的循环事件

现在我们要把循环事件的概念稍加推广．这个推广十分显然,除了为方便起见我们将给出有关的术语和基本方程外,其它方面的问题就不特别提及了．

粗糙地说,延迟了的循环事件是指它们所涉及的试验“不是从头开始而是从中间出发”，因此 $\mathscr{E}$ 第一次出现的等待时间的分布 $\{b_n\}$ 不同于 $\mathscr{E}$ 相继两次出现之间的循环时间的分布 $\{f_n\}$．在 $\mathscr{E}$ 出现后的试验是一个固定的样本空间的精确重复，但这个样本空间与原来的样本空间并非同一的．除此以外，前面的理论可以应用．

由于情况很简单，我们将不进行形式上的讨论，而作如下约定：如果当 $\mathscr{E}$ 第一次出现以前的各次试验不加考虑时，循环事件的定义仍可适用,且 $\mathscr{E}$ 第一次出现的等待时间是与以后的循环时间相独立的随机变量(前者的分布可以不同于共同的循环时间的分布 $\{f_n\}$)，则我们称 $\mathscr{E}$ 为延迟了的循环事件．

根据以上的定义及第 3 节中的结果容易计算出 $\mathscr{E}$ 在第 n 次试验中出现的概率 u_n．但更好的方法是写出一个更新型的新方程并独立地进行计算．

$\mathscr{E}$ 在第 $n-k$ 次试验中出现并且下一次在第 n 次试验中出现的概率为 $u_{n-k}f_k$. 这些事件是互不相容的，且它们对于 $k=1,2,\cdots,n-1$ 的并是如下事件：$\mathscr{E}$ 在第 n 次试验出现并且也在第 n 次试验前的某次试验中出现. $\mathscr{E}$ 在第 n 次试验第一次出现的概率为 b_n，故对于 $n\geqslant 1$ 有

$$u_n=b_n+u_{n-1}f_1+u_{n-2}f_2+\cdots+u_1f_{n-1}. \tag{5.1}$$

对于延迟了的事件，令

$$u_0=f_0=b_0=0 \tag{5.2}$$

是最自然的；这样就把 (5.1) 化归为更新方程

$$\{u_n\}=\{b_n\}+\{u_n\}*\{f_n\}, \tag{5.3}$$

并且对应的母函数满足

$$U(s)=\frac{B(s)}{1-F(s)}. \tag{5.4}$$

作为上节结果的一个特殊情况，我们有

定理. 如果 $\mathscr{E}$ 是非周期的，且 $\sum f_n=1$（即 $\mathscr{E}$ 是常返的），则

$$u_n\to\mu^{-1}\sum b_n,\ \ \mu=\sum nf_n. \tag{5.5}$$

如果 $f=\sum f_n<1$（即 $\mathscr{E}$ 是非常返的），则

$$\sum u_n=(1-f)^{-1}\sum b_n. \tag{5.6}$$

在周期的情况下，可以应用第 4 节的定理 2.

例. (a) 在计数器问题 (1.b) 中，假定在时刻 0 计数器正好关闭了两个时间单位（也就是说，在一次记录后经过两次试验才开始观察）. 计数至少还将关闭 $r-2$ 个时间单位，且如果第 $r-1$ 次试验的结果为失败，则计数器就在这次试验结束时成为自由的；否则它又将记录，因而继续关闭 r 个时间单位，等等. 由此推出

$$b_{r-2}=q,\ \ b_{2r-2}=pq,\ \ b_{3r-2}=p^2q,\ \ \cdots.$$

(b) **更新问题.** 在第 2 节的例题中，我们曾考虑过一种另件，其寿命是具有分布 $\{f_n\}$ 的随机变量. 当一个用坏了时，就立即用一个新的换上去，而且过程按这种方式继续下去. 设 $\mathscr{E}$ 代表"更换". 在第 2 节中我们假定，在时刻 0 正好安装上一个新另件. 现

在我们假定，在时刻 0，另件的年龄为 k. 这时 $\mathscr{E}$ 就是一个延迟了的循环事件，于是我们必须计算第一次更换的等待时间的分布 $\{b_n\}$. 显然 b_n 是已知另件的年龄为 k 的条件下另件的寿命为 $n+k$ 的概率. 因此有

$$b_n = \frac{f_{n+k}}{r_k}, \quad r_k = f_k + f_{k+1} + f_{k+2} + \cdots. \tag{5.7}$$

在应用中，通常不但要考虑单个另件，而且也要考虑整个总体. 设初始总体 (在时刻 0) 由 N 个另件所组成，其中年龄为 k 的恰有 v_k 个(显然 $\sum v_k = N$). 每个另件都引出一个逐代更替的另件系列，并在任意时刻 n，在这个系列中以某个概率需要更换. 对于所有 N 个另件的这些概率之和是在时刻 n 需要更换的期望数 u_n. 显然 u_n 满足基本方程 (5.3)，其中

$$b_n = \sum_{k=0}^{\infty} \frac{v_k f_{n+k}}{r_k}, \tag{5.8}$$

于是由前面的定理可知，u_n 是收敛的.

不仅 u_n 的极限，而且在时刻 n 的年龄分布及其渐近性质都是容易计算的. 设 $v_k(n)$ 是在时刻 n 年龄为 k 的另件的期望数(故有 $v_k(0) = v_k$). 显然有

$$\begin{aligned} v_k(n) &= u_{n-k} r_k, && \text{如果 } k < n; \\ v_k(n) &= \frac{v_{k-n} r_k}{r_{k-n}}, && \text{如果 } k \geqslant n. \end{aligned} \tag{5.9}$$

我们知道，在非周期的情况下，当 $n \to \infty$ 时，$u_n \to B(1)/\mu = N/\mu$，故由 (5.9) 可以推出 $v_k(n) \to N r_k/\mu$. 因此，在非周期的情况下，存在平稳的极限年龄分布，即：当 $n \to \infty$ 时，年龄为 k 的另件的期望数的极限为 $N r_k/\mu$，此处 N 是(不变的)总体的大小，$\mu = \sum r_k$ 为平均寿命(如果 $\mu = \infty$，则总体无限地老化). 基本事实是：极限年龄分布不依赖于初始年龄分布而仅与死亡分布 $\{a_n\}$ 有关(参见问题 17 与 18).

作为一个数值例题，我们考虑由 $N = 1000$ 个另件所组成的总体，其初始年龄分布为 $v_0 = 500$，$v_1 = 320$，$v_2 = 74$，$v_3 = 100$，

$v_4 = 6$. 设生存概率为 $f_1 = 0.20$, $f_2 = 0.43$, $f_3 = 0.17$, $f_4 = 0.17$, $f_5 = 0.03$（故最大年龄为 5）. 这里 $U(s)$ 是一个有理函数:

$$U(s) = s\frac{397 + 332s + 159s^2 + 97s^3 + 15s^4}{1 - 0.20s - 0.43s^2 - 0.17s^3 - 0.17s^4 - 0.03s^5}, \quad (5.10)$$

它能展开为部分分式:

$$U(s) = \frac{1250s}{3(1-s)} - \frac{972s}{61(1+3s/5)} + \frac{38s}{87(1+s/5)} - \frac{78225s^2 + 22125s}{5307(1+s^2/4)}.$$

年龄分布 $\{v_k(n)\}$ $(n = 1, 2, 3, \cdots)$ 可以直接由更新方程来计算. 表 1 各列给出了这些年龄分布及极限分布，它表明，$\{v_k(n)\}$ 在趋向其极限时并非单调的.

表 1

k	n								
	0	1	2	3	4	5	6	7	∞
0	500	397	411.4	412	423.8	414.3	417.0	416.0	416.7
1	320	400	317.6	329.1	329.6	339.0	331.5	333.6	333.3
2	74	148	185	146.9	152.2	152.4	156.8	153.3	154.2
3	100	40	80	100	79.4	82.3	82.4	84.8	83.3
4	6	15	6	12	15	11.9	12.3	12.4	12.5

(c) **人口问题.** 这个理论与更新理论类似，只是人口的多少是变化的，且女性的出生起着更换的作用. 这个问题的新颖之处在于，一个母亲可能有零个、一个或多个女儿，故谱系可能绝灭或分支. 现在我们定义 a_n 为一个新生的女孩成活并在年龄为 n 时生一个女孩的概率（对以前的子女的数目和年龄的依赖性略去不计）. 于是 $\sum a_n$ 是女儿的期望数，故 $\sum a_n < 1$, $\sum a_n = 1$, $\sum a_n > 1$ 都是可能的. 前面的论述，作一些明显的修改后，就可应用.

6. $\mathscr{E}$ 出现的次数

直到现在为止,我们仅考虑了一个循环事件的第一,第二,…,第 r 次出现,并把它们出现的试验数看作随机变量. 通常,更自然的是采取相反的观点,即固定试验数 n,并把在 n 次试验中 $\mathscr{E}$ 出现的次数 $\mathbf{N}_n$ 看作一个随机变量. 我们将研究当 n 很大时,$\mathbf{N}_n$ 的渐近性态.

如同在 (2.8) 中一样,设 $\mathbf{T}^{(r)}$ 表示第 r 次出现 $\mathscr{E}$ 的试验次数. $\mathbf{T}^{(r)}$ 与 $\mathbf{N}_n$ 的概率分布之间具有如下明显的关系:

$$\mathbf{P}\{\mathbf{N}_n \geqslant r\} = \mathbf{P}\{\mathbf{T}^{(r)} \leqslant n\}. \tag{6.1}$$

我们首先考虑 $\mathscr{E}$ 为常返的且其循环时间分布 $\{f_n\}$ 具有有限的均值 μ 与方差 σ^2 的简单情况. 因为 $\mathbf{T}^{(r)}$ 是 r 个独立随机变量之和,故由中心极限定理(第十章第 1 节)可以断定,对于每个固定的 x,当 $r \to \infty$ 时,有

$$\mathbf{P}\left\{\frac{\mathbf{T}^{(r)} - r\mu}{\sigma r^{1/2}} < x\right\} \to \Phi(x), \tag{6.2}$$

此处 $\Phi(x)$ 是正态分布函数. 令 $n \to \infty$ 和 $r \to \infty$,并使

$$\frac{n - r\mu}{\sigma r^{1/2}} \to x; \tag{6.3}$$

则由 (6.1) 与 (6.2) 可得

$$\mathbf{P}\{\mathbf{N}_n \geqslant r\} \to \Phi(x). \tag{6.4}$$

为了用更熟悉的形式写出这个关系,我们引进正则化随机变量

$$\mathbf{N}_n^* = \frac{\mathbf{N}_n - n\mu^{-1}}{\sigma n^{1/2}\mu^{-3/2}}. \tag{6.5}$$

不等式 $\mathbf{N}_n \geqslant r$ 可以写成如下形式

$$\mathbf{N}_n^* \geqslant \frac{r - n\mu^{-1}}{\sigma n^{1/2}\mu^{-3/2}} = -\frac{n - r\mu}{\sigma r^{1/2}} \cdot \left(\frac{r\mu}{n}\right)^{\frac{1}{2}}. \tag{6.6}$$

由 (6.3) 可推出,上式右边趋向于 $-x$,故

$$\mathbf{P}\{\mathbf{N}_n^* \geq -x\} \to \Phi(x) \text{ 或 } \mathbf{P}\{\mathbf{N}_n^* < -x\} \to 1 - \Phi(x), \quad (6.7)$$

于是我们证明了如下的

定理 1. 正态逼近. 如果循环事件 $\mathscr{E}$ 是常返的，且其循环时间具有有限的均值 μ 与方差 σ^2，则如 (6.2) 与 (6.7) 所示，第 r 次出现 $\mathscr{E}$ 的试验次数 $\mathbf{T}^{(r)}$ 与前 n 次试验中 $\mathscr{E}$ 出现的次数 $\mathbf{N}_n$ 都是渐近正态分布的.

注意，在 (6.7) 中，我们把中心极限定理应用到一列相依随机变量上去了. 由关系式 (6.7) 似乎应该有

$$\mathbf{E}(\mathbf{N}_n) \sim \frac{n}{\mu}, \quad \mathrm{Var}(\mathbf{N}_n) \sim \frac{n\sigma^2}{\mu^3}, \quad (6.8)$$

但其正确性尚需进一步论证.

下一节中我们将通过对连贯理论的应用来说明定理 1 的用处. 不过，应当了解，在随机变量起伏理论和物理过程中所遇到的循环时间大多数都有无限的均值，因而定理 1 必须用更一般的定理来代替[21].

在直观上，我们总以为 $\mathbf{E}(\mathbf{N}_n)$ 应当随着 n 线性地增大[如同在 (6.8) 中一样]，之所以产生这种想法，是由于如下的朴素考虑："试验的次数增加一倍时，$\mathscr{E}$ 出现的次数大体上也应增加一倍". 然而，事实并不如此. 掷钱币中的返回原点（它是扩散理论中的循环时间的代表）可以再次用来说明一般起伏问题中这种意料不到的性质.

定理 2. 循环悖理. 设 $\mathscr{E}$ 为对称伯努利试验（例如掷钱币）中的返回原点. 则在 $2n$ 次试验中 $\mathscr{E}$ 出现的期望次数 $\mathbf{E}(\mathbf{N}_{2n})$ 由下式给出

$$\mathbf{E}(\mathbf{N}_{2n}) = (2n+1)\binom{2n}{n}2^{-2n} - 1, \quad (6.9)$$

所以

$$\mathbf{E}(\mathbf{N}_{2n}) \sim 2(n/\pi)^{\frac{1}{2}}. \quad (6.10)$$

($\mathbf{E}(\mathbf{N}_n)$ 不是随 n 线性增加，而是按数量级 $n^{\frac{1}{2}}$ 增加.)

证. 由第十一章公式 (1.8) 及本章 (6.1) 可得

$$\mathbf{E}(\mathbf{N}_n)=\sum_{r=1}^{\infty}\mathbf{P}\{\mathbf{N}_n\geqslant r\}=\sum_{r=1}^{\infty}\mathbf{P}\{\mathbf{T}^{(r)}\leqslant n\}. \quad (6.11)$$

$\mathbf{T}^{(r)}$ 的母函数为 $F^r(s)$，其中 $F(s)$ 由 (3.16) 所决定，即 $F(s)=1-(1-s^2)^{\frac{1}{2}}$. 由第十一章第一节的定理 1 可知，累积概率 $\mathbf{P}\{\mathbf{T}^{(r)}\leqslant n\}$ 的母函数为 $F^r(s)(1-s)^{-1}$，于是由 (6.11) 可知，序列 $\{\mathbf{E}(\mathbf{N}_n)\}$ 的母函数为

$$\sum \mathbf{E}(\mathbf{N}_n)s^n=\frac{F(s)}{(1-s)(1-F(s))}$$

$$=\frac{1+s}{(1-s^2)^{\frac{3}{2}}}-\frac{1}{1-s}. \quad (6.12)$$

由此有

$$\mathbf{E}(\mathbf{N}_{2n})=\mathbf{E}(\mathbf{N}_{2n+1})=(-1)^n\binom{-\frac{3}{2}}{n}-1. \quad (6.13)$$

利用第二章公式 (12.5) 就可将 (6.13) 写成 (6.9) 的形式.

这个定理的奇妙含意我们曾在第三章第 6 节中详细讨论过. 那节的定理 2 证明，$\mathbf{N}_n\cdot n^{-2}$ 具有由正态分布的正值部分给出的渐近分布. 正则化变量 $\mathbf{N}_n n^{-2}$ 与以上定理 1 中的正则化变量形成鲜明的对比.

*7. 在成功连贯理论中的应用

在以下两节中，r 表示一个固定的正整数，$\mathscr{E}$ 代表在一列伯努利试验中长为 r 的成功连贯的出现. 连贯的长度按例 (1.a) 中的规定来定义是重要的，因为，不然的话，连贯就不是循环事件，因而计算就会变得更麻烦. 如在 (2.1) 与 (2.2) 中一样，u_n 是 $\mathscr{E}$ 在第 n 次试验中出现的概率，f_n 是长为 r 的成功连贯在第 n 次试验中第一次出现的概率.

---

* 本节及下节讨论专门的论题，可以略去.

第 $n, n-1, n-2, \cdots, n-r+1$ 次试验(共 r 次)都出现成功的概率显然为 p^r. 在这种情况下, $\mathscr{E}$ 在这 r 次试验之一中出现; $\mathscr{E}$ 在第 $n-k$ 次试验出现 $(k=0, 1, \cdots, r-1)$ 而在其后的 k 次试验都出现成功的概率为 $u_{n-k}p^k$. 因为这 r 种可能是互不相容的,故我们有如下的递推关系[1]:

$$u_n + u_{n-1}p + \cdots + u_{n-r+1}p^{r-1} = p^r. \tag{7.1}$$

这个方程当 $n \geqslant r$ 时成立. 显然

$$u_1 = u_2 = \cdots = u_{r-1} = 0, \quad u_0 = 1. \tag{7.2}$$

用 s^n 乘 (7.1) 并对 $n = r, r+1, r+2, \cdots$ 求和. 由 (7.2) 可知,左边的和为

$$\{U(s) - 1\}(1 + ps + p^2s^2 + \cdots + p^{r-1}s^{r-1}), \tag{7.3}$$

而右边的和则为 $p^r(s^r + s^{r+1} + \cdots)$. 求出这两个几何级数的和,我们得

$$\{U(s) - 1\} \cdot \frac{1-(ps)^r}{1-ps} = \frac{p^rs^r}{1-s} \tag{7.4}$$

或

$$U(s) = \frac{1 - s + qp^rs^{r+1}}{(1-s)(1-p^rs^r)}. \tag{7.5}$$

利用方程 (3.2), 我们得到循环时间的母函数

$$F(s) = \frac{p^rs^r(1-ps)}{1-s+qp^rs^{r+1}} = \frac{p^rs^r}{1-qs(1+ps+\cdots+p^{r-1}s^{r-1})}. \tag{7.6}$$

因为 $F(1) = 1$, 故在延长的试验序列中, 任意长的连贯数必定无限增加. 因为 $u_n \to \mu^{-1}$, 故平均循环时间可直接由 (7.1) 得到. 如果还要求方差,则用 $F(s)$ 的导数来求更好一些. 在求 $F'(s)$ 时, 最好是先去 (7.6) 的分母, 然后用隐函数微分法. 求出 $F'(s)$ 后,再通过简单的计算就可证明,长为 r 的连贯的循环时间的均值

1) 古典的方法在于导出 f_n 的递推关系. 这个方法要麻烦得多, 因而不能应用于任一种连贯或像 SSFFSS 之类的型样; 而我们的方法,无需改变,就可用来讨论这类问题.

与方差分别为

$$\mu=\frac{1-p^r}{qp^r},\quad \sigma^2=\frac{1}{(qp^r)^2}-\frac{2r+1}{qp^r}-\frac{p}{q^2}. \tag{7.7}$$

由上节定理 1 可知，当 n 很大时，n 次试验中所产生的长为 r 的连贯数 $\mathbf{N}_n$ 近似地服从正态分布，即对于固定的 $\alpha<\beta$，

$$\frac{n}{\mu}+\frac{\alpha\sigma n^{\frac{1}{2}}}{\mu^{\frac{3}{2}}}<\mathbf{N}_n<\frac{n}{\mu}+\frac{\beta\sigma n^{\frac{1}{2}}}{\mu^{\frac{3}{2}}} \tag{7.8}$$

的概率趋向于 $\Phi(\beta)-\Phi(\alpha)$. 这个事实首先是由冯·米赛斯证明的，但如没有循环事件的理论，则证明需要相当冗长的计算. 表 2 给出了几个典型的循环时间的平均值.

表 2

成功连贯的平均循环时间．假定试验以每秒一次的速率进行．

连贯的长度	$p=0.6$	$p=0.5$（钱币）	$p=\frac{1}{6}$（骰子）
$r=5$	30.7 秒	1 分	2.6 小时
10	6.9 分	34.1 分	28.0 月
15	1.5 小时	18.2 小时	18098 年
20	19 小时	24.3 天	140.7 百万年

利用第九章第 4 节中的部分分式的方法可以导出一个极好的近似公式. 容易看出，(7.6)中第二个表示式的分母有唯一的正根 $s=x$. 对于满足 $|s|\leqslant x$ 的每个实数或虚数 s，我们有

$$\begin{aligned}&|qs(1+ps+\cdots+p^{r-1}s^{r-1})|\\&\qquad\leqslant qx(1+px+\cdots+p^{r-1}x^{r-1})=1,\end{aligned} \tag{7.9}$$

其中等号仅当左端的所有项具有相同的幅角（即 $s=x$）时才能成立. 故 x 的绝对值小于 (7.6) 的分母的任何其它根的绝对值. 因此我们可以应用第十一章中的公式 (4.5) 与 (4.9)，其中 $s_1=x$. 系数 ρ_1 容易由 $U(s)=p^rs^r(1-ps)$ 及 $V(s)=1-s+qp^rs^{r+1}$ 算出. 利用 $V(x)=0$ 即可求得

$$f_n\sim\frac{(x-1)(1-px)}{(r+1-rx)q}\cdot\frac{1}{x^{n+1}}. \tag{7.10}$$

在 n 次试验中无连贯的概率为 $q_n = f_{n+1} + f_{n+2} + f_{n+3} + \cdots$. 由 (7.10) 可知,可用一个几何级数来近似 q_n, 于是我们得到

$$q_n \sim \frac{1 - px}{(r + 1 - rx)q} \cdot \frac{1}{x^{n+1}}. \tag{7.11}$$

故在 n 次试验中不出现长为 r 的成功连贯的概率渐近地由 (7.11) 给出. 表 3 表明, 甚至当 n 很小时, 公式 (7.11) 给出的近似也极为精确, 且近似程度随着 n 的增加而迅速改进. 这表明, 母函数和部分分式的方法是一种有力的工具.

表 3

$p = \frac{1}{2}$ 的 n 次试验中不出现长为 $r = 2$ 的成功连贯的概率.

n	q_n 的精确值	(7.11) 给出的近似	误差
2	0.75	0.76631	0.0163
3	0.625	0.61996	0.0080
4	0.500	0.50156	0.0016
5	0.40625	0.40577	0.0005

数值计算. 为了对注重实用的读者有所裨益,我们借此机会来说明,部分分式展开中所牵涉的数值计算经常并不像乍一看来那样繁难,并且能得到极好的误差估计.

渐近展式 (7.11) 提出两个问题: 首先, 必须估计略去的 $r - 1$ 个根的影响;其次,必须求出起主要作用的根 x.

(7.6) 中的第一式表明, $F(s)$ 的分母的所有根满足方程

$$s = 1 + qp^r s^{r+1}. \tag{7.12}$$

易知, $s = p^{-1}$ 是 (7.12) 的一个根. 对于正的 s, $f(s) = 1 + qp^r s^{r+1}$ 是凸的;它在 x 与 p^{-1} 处与平分线 $y = s$ 相交,故在 x 与 p^{-1} 之间其图形在平分线的下方. 此外, $f'(p^{-1}) = (r + 1)q$. 如果这个数大于 1, 则 $f(s)$ 的图形从下面穿过平分线,因而 $p^{-1} > x$. 为确定起见,我们假定

$$(r + 1)q > 1; \tag{7.13}$$

在这种情况下, $x < p^{-1}$, 且当 $x < s < p^{-1}$ 时, $f(s) < s$. 由此可以推出,对于满足不等式 $x < |s| < p^{-1}$ 的所有复数 s, 我们有 $|f(s)| \leq f(|s|) < |s|$, 故任何根 s_k 都不可能位于环形区域 $x < |s| < p^{-1}$ 中. 因为 x 是绝对值最小的根,故对于每个根 $s_k \neq x$ 有

$$|s_k| > p^{-1}. \tag{7.14}$$

通过求 (7.12) 的导数可以看出，所有根都是单根。

q_n 的相应于每个根的部分与它相应于起主要作用的根 x 的部分 (7.11) 具有相同的形式，故 (7.11) 中所略去的 $r-1$ 项具有如下形式

$$A_k = \frac{ps_k - 1}{rs_k - (r+1)} \cdot \frac{1}{q\, s_k^{n+1}}. \tag{7.15}$$

我们要估计上式右边第一个分式的上界。为此我们注意，对于固定的 $s > p^{-1} > (r+1)r^{-1}$ 有

$$\left|\frac{pse^{i\theta} - 1}{rse^{i\theta} - (r+1)}\right| \leqslant \frac{ps+1}{rs+r+1}; \tag{7.16}$$

事实上，上式左边显然当 $\theta = 0$ 和 $\theta = \pi$ 时取极值，以 0 与 π 直接代入即可看出，$\theta = 0$ 对应于极小值，$\theta = \pi$ 对应于极大值。由 (7.13) 与 (7.14) 有

$$|A_k| < \frac{2p^{n+1}}{(r+1+rp^{-1})q} < \frac{2p^{n+2}}{rq(1+p)}. \tag{7.17}$$

于是我们得到如下结论：在 (7.11) 中由于略去 $r-1$ 个不同于 x 的根而引起的误差的绝对值小于

$$\frac{2(r-1)p}{rq(1+p)}. \tag{7.18}$$

根 x 容易用逐次逼近法来计算 [令 $x_0 = 1$, $x_{\nu+1} = f(x_\nu)$]。所得出的序列单调收敛于 x，因而每一项都可作为 x 的一个下界，而使得 $s > f(s)$ 的任何值 s 都可作为一个上界。容易看出，

$$x = 1 + qp^r + (r+1)(qp^r)^2 + \cdots. \tag{7.19}$$

8. 更一般的型样

我们的方法可以应用于一类比连贯理论深奥得多的更为一般的问题。

例． (a) **不限种类的连贯**．设 $\mathscr{E}$ 代表“一个长 r 的成功连贯或一个长为 ρ 的失败连贯．” 这里我们所涉及的是两个循环事件 $\mathscr{E}_1$ 与 $\mathscr{E}_2$，其中 $\mathscr{E}_1$ 代表“长为 r 的成功连贯”，$\mathscr{E}_2$ 代表“长为 ρ 的失败连贯”，而 $\mathscr{E}$ 则为 “$\mathscr{E}_1$ 或 $\mathscr{E}_2$”．对应于 $\mathscr{E}_1$ 的母函数为 (7.5)，现在我们将它记为 $U_1(s)$．对应于 $\mathscr{E}_2$ 的母函数可由在 (7.5) 中交

换 p 与 q 并用 ρ 代替 r 而得. $\mathscr{E}$ 在第 n 次试验中出现的概率 u_n，除 u_0 外，是对于 $\mathscr{E}_1$ 与 $\mathscr{E}_2$ 的相应概率之和. 由此有

$$U(s) = U_1(s) + U_2(s) - 1. \tag{8.1}$$

于是 $\mathscr{E}$ 的循环时间的母函数 $F(s)$ 为 $F(s) = 1 - U^{-1}(s)$ 或

$$F(s) = \frac{(1-ps)p^r s^r(1-q^\rho s^\rho) + (1-qs)q^\rho s^\rho(1-p^r s^r)}{1 - s + qp^r s^{r+1} + pq^\rho s^{\rho+1} - p^r q^\rho s^{r+\rho}}. \tag{8.2}$$

由微分法可求得平均循环时间

$$\mu = \frac{(1-p^r)(1-q^\rho)}{qp^r + pq^\rho - p^r q^\rho}. \tag{8.3}$$

当 $\rho \to \infty$ 时，上式趋向于由 (7.7) 给出的成功连贯的平均循环时间.

(b) 在第八章第 1 节中，我们计算过长为 r 的成功连贯发生在长为 ρ 的失败连贯之前的概率. 如例 (a) 中那样定义两个循环事件 $\mathscr{E}_1$ 与 $\mathscr{E}_2$. 设 x_n 为 $\mathscr{E}_1$ 在第 n 次试验第一次出现且在此之前 $\mathscr{E}_2$ 未曾出现的概率; f_n 为 $\mathscr{E}_1$ 在第 n 次试验第一次出现的概率(对于 $\mathscr{E}_2$ 不加任何限制). 交换 $\mathscr{E}_1$ 与 $\mathscr{E}_2$，像定义 x_n 与 f_n 那样，分别给出 y_n 与 g_n 的定义.

在 (7.6) 中给出了 f_n 的母函数，交换 p 与 q 并用 ρ 代替 r 即得 $G(s)$. 对于 x_n 与 y_n，显然有如下的递推关系:

$$\begin{aligned} x_n &= f_n - (y_1 f_{n-1} + y_2 f_{n-2} + \cdots + y_{n-1} f_1), \\ y_n &= g_n - (x_1 g_{n-1} + x_2 g_{n-2} + \cdots + x_{n-1} g_1). \end{aligned} \tag{8.4}$$

这些方程是褶积型的，故对于对应的母函数我们有

$$\begin{aligned} X(s) &= F(s) - Y(s)F(s), \\ Y(s) &= G(s) - X(s)G(s). \end{aligned} \tag{8.5}$$

解这两个线性方程可得

$$X(s) = \frac{F(s)\{1 - G(s)\}}{1 - F(s)G(s)}, \quad Y(s) = \frac{G(s)\{1 - F(s)\}}{1 - F(s)G(s)}. \tag{8.6}$$

x_n 与 y_n 的表达式又可由部分分式的方法求得. 当 $s = 1$ 时，我们得 $X(1) = \sum x_n = x$，这就是 $\mathscr{E}_1$ 在 $\mathscr{E}_2$ 之前出现的概率. 当 $s = 1$ 时，(8.6) 第一式的分子、分母都为 0，故 $X(1)$ 可由洛必大法则通

过求分子、分母的导数而得到：

$$X(1) = G'(1)/\{F'(1) + G'(1)\}.$$

由 (7.7) 有

$$F'(1) = (1 - p^r)/qp^r,\quad G'(1) = (1 - q^\rho)/pq^\rho.$$

以此代入上式即得 $X(1)$，它与第八章公式 (1.3) 是一致的.

(c) 考虑由型样 $SSFFSS$ 所定义的循环事件. 重复第 7 节中的论述，容易求得

$$p^4q^2 = u_n + p^2q^2u_{n-4} + p^3q^2u_{n-5}. \tag{8.7}$$

因为 $u_n \to \mu^{-1}$，故由上式即可求出平均循环时间 $\mu = p^{-4}q^{-2} + p^{-2} + p^{-1}$. 当 $p = q = 1/2$ 时有 $\mu = 70$，而长为 6 的成功连贯的平均循环时间为 126. 这说明，出乎意料地，在掷钱币中的成功连贯与具有相同长度的其它连贯之间有着本质的区别.

9. 几何等待时间的记忆缺乏

等待时间的几何分布有一个任何其它分布都不具备的重要而有趣的性质. 考虑一列伯努利试验，并设 $\mathbf{T}$ 为直到第一次出现成功所需的试验(包括出现成功那次试验在内)的次数，设 $\mathbf{P}\{\mathbf{T} > k\} = q^k$. 假定我们已知在前 m 次试验中没有出现成功；则从这第 m 次失败起一直到第一次成功的等待时间也具有同一分布 $\{q^k\}$，且它与前面的失败次数相独立. 换句话说，等待时间延长 k 的概率和等待的总长超过 k 的初始概率是相等的. 如果一个原子或某个另件的寿命具有几何分布，则不会发生老化现象；只要原子还存在，它在下一次试验中衰变的概率总是一样的. 实际上，放射性原子就具有这种性质(只是在时间连续的情况下，指数分布起着几何分布的作用). 反之，如果一种现象具有完全缺乏记忆(或不会老化)的特点，则它的持续时间必定服从几何分布或指数分布. 一个典型的例子是电话的会话(它常被用来作为前后不连贯和完全依赖于瞬时冲动的模型)，它的可能结束是一个瞬时的偶然结果，而与结束前会话的长短无关. 与此相反，如果已有 5 分钟没有公共汽

车经过，则它马上就要来的可能性就更大．在掷钱币时，正面和反面的累积数相等在第二次试验出现的概率为 1/2．然而，在已知它们不相等的条件下，再经过两次试验而出现相等的概率仅仅是 1/4．这些都是有后效的例子．

现在我们来严视地系统阐述上面的断言．假定等待时间 $\mathbf{T}$ 分别以概率 $p_0, p_1, p_2, \cdots$ 取值 $0, 1, 2, \cdots$．设 $\mathbf{T}$ 的分布具有如下性质：在第 k 次试验之前等待时间尚未结束的条件下，它在第 k 次试验结束的条件概率等于 p_0（即在第一次试验结束的概率），则 $p_k = (1-p_0)^k p_0$，即 $\mathbf{T}$ 服从几何分布．

证．令

$$q_k = p_{k+1} + p_{k+2} + p_{k+3} + \cdots = \mathbf{P}\{\mathbf{T} > k\}.$$

我们的条件为 $\mathbf{T} > k-1$，其概率为 q_{k-1}．因此 $\mathbf{T} = k$ 的条件概率为 p_k/q_{k-1}，于是以上的假设可以表为，对于所有的 $k \geqslant 1$ 有

$$\frac{p_k}{q_{k-1}} = p_0. \tag{9.1}$$

因为 $p_k = q_{k-1} - q_k$，故

$$\frac{q_k}{q_{k-1}} = 1 - p_0. \tag{9.2}$$

因为 $q_0 = p_1 + p_2 + \cdots = (1-p_0)^k p_0$，故有 $q_k = (1-p_0)^{k+1}$，因而 $p_k = q_{k-1} - q_k = (1-p_0)^k p_0$，这就是所要证明的．

在随机过程理论中，以上所叙述的记忆缺乏是和马尔可夫性质相联系的；我们将在第十五章第 10 节中再来讨论这个问题．

*10. 第 3 节定理 3 的证明

在第 3 节中我们略去了定理 3 的证明．这个定理可以作为关于幂级数的陶贝尔（Tauber）定理的一个特例，也可用如下的初等方法来证明．设 $\{f_n\}$ 是一已给数列，其中 $f_0 = 0$，$f_n \geqslant 0$，

---

* 初次阅读时，本节可以略去．

$\sum f_n = 1$，且使得 $f_n > 0$ 的那些 n 的最大公约数为 1．设 $u_0 = 1$，当 $n \geqslant 1$ 时，u_n 由下式定义

$$u_n = f_1 u_{n-1} + f_2 u_{n-2} + \cdots + f_n u_0, \tag{10.1}$$

则 $u_n \to 1/\mu$，此处 $\mu = \sum n f_n$（如果 $\sum n f_n$ 发散，则 $u_n \to 0$）．

证．令

$$r_n = f_{n+1} + f_{n+2} + \cdots, \tag{10.2}$$

则由第十一章公式 (1.8) 有

$$\mu = \sum r_n. \tag{10.3}$$

由 (10.2) 可得 $r_0 = 1$，$f_1 = r_0 - r_1$，$f_2 = r_1 - r_2$，等等．以这些值代 (10.1)，我们得

$$r_0 u_n + r_1 u_{n-1} + \cdots + r_n u_0 = r_0 u_{n-1} + r_1 u_{n-2} + \cdots + r_{n-1} u_0.$$

如果用 A_n 表示上式的左边，则上式的右边为 A_{n-1}，而由这个等式可知，所有的 A_n 都是相等的．因为 $A_0 = r_0 u_0 = 1$，故对一切 n 有 $A_n = 1$．于是对每个 n 我们有

$$r_0 u_n + r_1 u_{n-1} + \cdots + r_n u_0 = 1. \tag{10.4}$$

由(10.1)用归纳法可以推出 $u_n \leqslant 1$．故存在数 $\lambda = \limsup u_n$，使得对于任意的 $\varepsilon > 0$ 及所有充分大的 n 有 $u_n < \lambda + \varepsilon$，且存在某个序列 $n_1, n_2, n_3, \cdots$，使得 $u_{n_\nu} \to \lambda$．选取正整数 j，使得 $f_j > 0$．我们要证明 $u_{n_\nu - j} \to \lambda$．如果不然，则可以找到任意大的足标 n，使下面两式同时成立：

$$u_n > \lambda - \varepsilon, \quad u_{n-j} < \lambda' < \lambda. \tag{10.5}$$

现在设 N 充分大，使得 $r_N < \varepsilon$．因为 $u_k \leqslant 1$，故由 (10.1)，当 $n > N$ 时有

$$u_n \leqslant f_0 u_n + f_1 u_{n-1} + \cdots + f_N u_{n-N} + \varepsilon. \tag{10.6}$$

当 n 充分大时，上式右边的每个 u_k 都小于 $\lambda + \varepsilon$，且 $u_{n-j} < \lambda'$．故有

$$\begin{aligned} u_n &< (f_0 + f_1 + \cdots + f_{j-1} + f_{j+1} + \cdots + f_N)(\lambda + \varepsilon) \\ &\quad + f_j \lambda' + \varepsilon \leqslant (1 - f_j)(\lambda + \varepsilon) + f_j \lambda' + \varepsilon \\ &< \lambda + 2\varepsilon - f_j(\lambda - \lambda'). \end{aligned} \tag{10.7}$$

如果我们取 ε 充分小，使得 $f_j(\lambda - \lambda') > 3\varepsilon$，则这个不等式与

(10.5) 中的第一个不等式矛盾，故 $\lambda' < \lambda$ 的假设是不可能的.

这就证明了，如果 $u_{n_\nu} \to \lambda$，则 $u_{n_\nu - j} \to \lambda$. 重复以上论证，我们有：如果 $f_j > 0$ 且 $u_{n_\nu} \to \lambda = \limsup u_n$，则也有

$$u_{n_\nu - j} \to \lambda, \quad u_{n_\nu - 2j} \to \lambda, \quad u_{n_\nu - 3j} \to \lambda, \quad \text{等等}.$$

为简单起见，我们首先考虑 $f_1 > 0$ 的情况. 于是我们可取 $j = 1$，并推出，对于每个固定的 k 有 $u_{n_\nu - k} \to \lambda$. 在 (10.4) 中令 $n = n_\nu$ 可得

$$1 \geqslant r_0 u_{n_\nu} + r_1 u_{n_\nu - 1} + \cdots + r_N u_{n_\nu - N}. \tag{10.8}$$

对于固定的 N，每个 $u_{n_\nu - k} \to \lambda$，故

$$1 \geqslant \lambda(r_0 + r_1 + \cdots + r_N).$$

因为 N 是任意的，故有 $1 > \lambda\mu$ 或 $\lambda < 1/\mu$. 于是，如果 (10.3) 发散，则有 $u_n \to 0$，故定理得证.

如果 $\mu < \infty$，则令 $\gamma = \liminf u_n$. 用同样的方法可以证明，对于每个序列 n_ν，如果 $u_{n_\nu} \to \gamma$，则 $u_{n_\nu - k} \to \gamma$ 也成立. 如果 N 足够大，使得 $r_N < \varepsilon$，则由 (10.4) 有

$$1 \leqslant r_0 u_{n_\nu} + \cdots + r_N u_{n_\nu - N} + \varepsilon; \tag{10.9}$$

因为 $u_{n_\nu - k} \to \gamma$，故 $1 \leqslant (r_0 + \cdots + r_N)\gamma + \varepsilon$，因此 $\mu\gamma \geqslant 1$. 然而，根据定义，$\gamma \leqslant \lambda$，所以 $\gamma = \lambda = 1/\mu$. 证毕.

现在讨论 $f_1 = 0$ 的情况. 考虑使 $f_i > 0$ 的所有整数 i 的集合. 在它们中间能找到一个其最大公约数为 1 的有限集 $a, b, c, \cdots, m$. 我们知道，当 $u_{n_\nu} \to \lambda$ 时，对于每个固定的 $x > 0, y > 0, \cdots, w > 0$，也有 $u_{n_\nu - xa} \to \lambda$，$u_{n_\nu - yb} \to \lambda$，等等；因此也有

$$u_{n_\nu - xa - yb - \cdots - wm} \to \lambda.$$

换句话说，如果整数 k 能表成 $k = xa + yb + \cdots + wm$ 的形式（其中 $x, y, \cdots, w$ 为正整数），则 $u_{n_\nu - k} \to \lambda$. 由初等数论可知，每个大于乘积 $abc\cdots m$ 的整数都能写成这种形式. 这意味着，对于 $k > abc\cdots m$ 我们有 $u_{n_\nu - k} \to \lambda$. 在 (10.4) 中令 $n = n_\nu + abc\cdots m$ 就可以得到 (10.8). 证明的其余部分无需改变.

11. 问　　题

1. 设 $F(s)$ 为一多项式．用第十一章第 4 节中的部分分式方法证明，在这种情况第 3 节中的所有定理成立．

2. 重复掷 r 个钱币，设 $\mathscr{E}$ 表示如下的循环事件：对于 r 个钱币中的每一个出现的正面和反面的累积数相等．$\mathscr{E}$ 是常返的还是非常返的？对于使 $\mathscr{E}$ 为非常返的最小 r，估计 $\mathscr{E}$ 迟早要出现的概率．

3. 设 $\{\mathbf{X}_k\}$ 为一列相互独立的随机变量，它们具有共同的分布 $\mathbf{P}\{\mathbf{X}_k = a\} = b/(a+b)$，$\mathbf{P}\{\mathbf{X}_k = -b\} = a/(a+b)$，其中 a 与 b 为正整数．设 $\mathscr{E}$ 表示事件 $\mathbf{S}_n = 0$．证明 $\mathscr{E}$ 是常返事件．

4. 设 $\{\mathbf{X}_k\}$ 是任意一列独立同分布的随机变量，又设 $\mathscr{E}$ 代表 $\mathbf{S}_n = 0$，$\mathbf{S}_1 \leqslant 0$，$\mathbf{S}_2 \leqslant 0$，$\cdots$，$\mathbf{S}_{n-1} \leqslant 0$．证明除 $\mathbf{P}\{\mathbf{X}_k = 0\} = 1$ 这个特殊情况外，$\mathscr{E}$ 是一个非常返的循环事件．

5. **累次平均**．将第 4 节中的例子推广到加权平均值上去，并求出极限．

注意．问题 6—8 参照 $p = q = 1/2$ 的伯努利试验（掷钱币）．假定返回原点的母函数 $F(s) = 1 - (1 - s^2)^{1/2}$ 是已知的．

6. 设 $\mathscr{E}_1$ 为循环事件 $\mathbf{S}_n = 0$，$\mathbf{S}_{n-1} < 0$．求循环时间的母函数．

7. （续上）．求梯级点（例 1.d）的循环时间的母函数．（注意．这与第十一章第 3 节中所讨论的初过 1 的等待时间是一样的．）

8. （续上）．证明定理：$\mathbf{S}_k$ 在第 $2n$ 或第 $2n+1$ 次试验中取到它以前所未曾取到过的值（即出现一个初过）的概率等于 $\mathbf{S}_{2n} = 0$ 的概率．

9. 在计数器问题（1.b）中：(a) 求循环时间的母函数．（其物理意义是什么？）(b) 如果 $\mathbf{Z}_n$ 为前 n 次试验中的记录数，求 $\mathbf{E}(\mathbf{Z}_n)$ 与 $\mathrm{Var}(\mathbf{Z}_n)$．

10. 第二类计数器与例（1.b）中的计数器不同之处在于，每次成功使计数器关闭 r 个时间单位（出现成功后的 $r-1$ 次试验），且关闭期间内的成功会使关闭的时间延长．对这种计数器解问题 9．

11. 求在钱币的 10000 次投掷中长为 3 的正面连贯数在 700 与 730 之间的概率的近似值．

12. 在掷钱币的序列中，设 $\mathscr{E}$ 代表型样 HTH，r_n 为 $\mathscr{E}$ 在第 n 次试验中不出现的概率．求母函数，并用部分分式的方法求出渐近展式．

13. 证明在例（8.b）中赌博的持续时间的期望值为

$$\mu_1\mu_2/(\mu_1 + \mu_2),$$

其中 μ_r 与 μ_ρ 分别为长为 r 的成功连贯与长为 ρ 的失败连贯的平均循环时间。

14. 设每次试验的可能结果为 A, B, C; 对应的概率为 α, β, γ $(\alpha+\beta+\gamma=1)$. 求在 n 次试验中不出现如下的长为 r 的连贯的概率的母函数: (a) A 的连贯, (b) A 或 B 的连贯. (c) 任一种连贯.

15. (续上). 求第一个长为 r 的 A-连贯在第 n 次试验中出现且在此之前未曾出现过长为 ρ 的 B-连贯的概率. [注意, 这个问题不能化归为当 $p=\alpha/(\alpha+\beta)$, $q=\beta/(\alpha+\beta)$ 时的例 (8.b) 中的问题.]

注意. 下面的问题参照更新理论,特别是例 (5.b).

16. 总体大小的不变性. 对于 (5.9) 中的量,用归纳法证明对于每个 n 有 $\sum_k v_k(n)=N$.

17. 设死亡分布为 $p_k=q^{k-1}p(p+q=1)$, 且初始总体由 N 个年龄为 0 的元素组成,求 u_n 及极限年龄分布.

18. 如果 $v_k(n)$ 不依赖于 n, 则称年龄分布是平稳的. 证明当且仅当 $v_k=Cr_k$(C 为常数)时,年龄分布是平稳的.

---

19. 设 $\mathscr{E}$ 为常返的非周期循环事件. 假定循环时间具有有限的平均值 μ 和方差 σ^2. 令 $q_n=f_{n+1}+f_{n+2}+\cdots$ 及 $r_n=q_{n+1}+q_{n+2}+\cdots$. 证明母函数 $Q(s)$ 与 $R(s)$ 当 $s=1$ 时收敛. 证明

$$\sum\left(u_n-\frac{1}{\mu}\right)s^n=\frac{R(s)}{\mu Q(s)}, \tag{11.1}$$

因而证明

$$\sum\left(u_n-\frac{1}{\mu}\right)=\frac{\sigma^2-\mu+\mu^2}{2\mu^2}. \tag{11.2}$$

20. 设 $\mathscr{E}$ 为常返循环事件, $\mathbf{N}_r$ 为在 r 次试验中 $\mathscr{E}$ 出现的次数. 证明 $\mathbf{E}(\mathbf{N}_r)=u_1+\cdots+u_n$, 因而

$$\mathbf{E}(\mathbf{N}_r)\sim\frac{r}{\mu}. \tag{11.3}$$

21. (续上). 证明

$$\mathbf{E}(\mathbf{N}_r^2)=u_1+\cdots+u_r+2\sum_{j=1}^{r-1}u_j(u_1+\cdots+u_{r-j})$$

因而 $\mathbf{E}(\mathbf{N}_r^2)$ 是

$$\frac{F^2(s)+F(s)}{(1-s)\{1-F(s)\}^2} \tag{11.4}$$

中 s^r 的系数。(注意：这也可以更巧妙地重新用二元母函数来表达.)

22．设 $q_{k,n}=\mathbf{P}\{\mathbf{N}_k=n\}$． 证明 $q_{k,n}$ 是

$$F^n(s)\frac{\{1-F(s)\}}{1-s} \tag{11.5}$$

中 s^k 的系数。证明 $\mathbf{E}(\mathbf{N}_r)$ 与 $\mathbf{E}(\mathbf{N}_r^2)$ 分别为

$$\frac{F(s)}{(1-s)\{1-F(s)\}} \tag{11.6}$$

与 (11.4) 中 s^r 的系数。

23．使用问题 19 中的记号，证明

$$\frac{F(s)}{(1-s)\{1-F(s)\}}=-\frac{1}{1-s}+\frac{1}{\mu(1-s)^2}+\frac{R(s)}{\mu\{1-F(s)\}}. \tag{11.7}$$

因此由上一个问题可以推出

$$\mathbf{E}(\mathbf{N}_r)=\frac{r+1}{\mu}+\frac{\sigma^2-\mu-\mu^2}{2\mu^2}+\varepsilon_r, \tag{11.8}$$

其中 $\varepsilon_r\to 0$。

24．(续上)．用类似的方法证明

$$\mathbf{E}(\mathbf{N}_r^2)=\frac{(r+2)(r+1)}{\mu^2}+\frac{2\sigma^2-2\mu-\mu^2}{\mu^3}r+\alpha_r, \tag{11.9}$$

其中 α_r 为有界。因此

$$\mathrm{Var}(\mathbf{N}_r)\sim\frac{\sigma^2}{\mu^3}r. \tag{11.10}$$

25．在一列伯努利试验中，令 $q_{k,n}$ 表示在 k 次试验中恰有 n 个长为 r 的成功连贯出现的概率．利用问题 22 证明母函数 $Q_k(x)=\sum q_{k,n}x^n$ 是

$$\frac{1-p^rs^r}{1-s+qp^rs^{r+1}-(1-ps)p^rs^rx}$$

中 s^k 的系数。而且证明，分母的绝对值最小的根为 $s_1\approx 1+qp^r(1-x)$。

26．(续上)．**长连贯的普阿松分布**[1]．如果试验数 k 与连贯长度 r 两者都趋向于无穷，且使得 $kqp^r\to\lambda$，则恰有 n 个长为 r 的连贯的概率趋向于 $e^{-\lambda}\lambda^n/n!$．

提示：利用前面的问题，证明母函数渐近地为 $\{1+qp^r(1-x)\}^{-k}\sim e^{-\lambda(1-x)}$．利用第十一章第 6 节中的连续性定理。

---

1) 这个定理是冯·米赛斯证明的，但现在的方法要简单得多。

第十四章　随机徘徊与输光问题

1. 一般讨论

本章第一部分讨论伯努利试验，为使叙述生动，我们再次使用赌博和随机徘徊的形象语言.

考虑在每次试验中分别以概率 p 与 q 赢一元或输一元的赌徒. 设他的初始赌本为 z，对手的初始赌本为 $a-z$，故联合赌本为 a. 赌博一直进行到赌徒的赌本减少到 0 或增加到 a 为止，也就是说，进行到两赌徒有一个输光为止，我们所关心的是赌徒输光的概率及赌博持续时间的概率分布. 这就是古典输光问题.

物理中的种种应用与模拟促使我们采用变点或"质点"在 x 轴上的运动这种更加易于掌握的解释. 在时刻 0 质点位于初始位置 z，在每一时刻 1, 2, 3, …, 如果对应的试验结果为成功，则它沿 x 轴的正向移动一个单位步；如果试验的结果为失败，则它沿 x 轴的负向移动一个单位步. 质点在时刻 n 的位置表示在第 n 次试验结束时赌徒的赌本. 当质点第一次到达 0 或 a 时试验结束. 我们称质点的这种运动为在 0 与 a 处具有吸引壁的随机徘徊. 这种随机徘徊被限制在 $1, 2, \cdots, a-1$ 等可能位置；如果没有吸收壁，则称随机徘徊是自由的. 物理学家用随机徘徊作为一维扩散或布朗运动的一个粗糙的近似，其中物理质点由于受到大量分子的碰撞而产生随机运动. $p>q$ 的情况对应于右向趋向，这时质点受到来自左方的冲击的可能性较大；如果 $p=q=1/2$，则称随机徘徊为对称的.

在 $a\to\infty$ 的极限情况，我们得到在半无限直线上的随机徘徊. 质点从 $z>0$ 出发作随机徘徊，直到它第一次达到原点时为止. 在问题的这种陈述中，我们可以看出，这是一个初过时间的问

题；在第三章中我们曾给出它的初等解法（至少就对称的情况），在第十一章第 3 节中又利用母函数的工具给出了它的另一种解法（又见第十三章问题 7）. 我们将看到，本章中有些公式是以前曾得到过的，不过，现在的推导方法是新的[1].

在本章中我们将使用差分方程的方法，这个方法可作为扩散理论的微分方程的一个导引. 这种模拟自然地导致古典输光问题的各种修改和推广，一个典型的、富有教益的例子是用反射壁和弹性壁来代替吸收壁. 为了描述反射壁，考虑在区间 $(0, a)$ 中的一个随机徘徊，除作如下修改外，它的定义和前面一样：当质点位于点 1 时，它以概率 1 运动到位置 2，而以概率 q 留在 1. 用赌博的话来说，这对应于这样一种规定：当赌徒输掉他最后的一元时，他的对手慷慨地把这一元还给他，因而赌博可以继续下去. 物理学家想像在 x 轴的点 1/2 处有具有如下性质的一面墙：由 1 向 0 运动的质点都被墙反射，故质点回到 1 而不能到达 0. 吸收壁和反射壁都是所谓弹性壁的特殊情况. 我们用如下规则来定义在原点处的弹性壁：质点从位置 1 以概率 p 运动到 2；以概率 δq 停留在 1 处；以概率 $(1-\delta)q$ 运动到 0 而被吸收（即过程结束）. 当 $\delta=0$ 时我们就得到古典输光问题或吸收壁，当 $\delta=1$ 时则得到反射壁. 当 δ 由 0 变到 1 时我们就得到一系列中间情况. δ 越大，过程持续下去的可能性就越大，如果有两个反射壁，则过程永远不会结束.

在第 2、3 两节中我们将给出古典输光问题及其推论的初等讨论. 其后三节的内容比较专门（因而可以略去）；在第 4、5 两节中我们导出相关母函数并由此导出赌博持续时间分布的明显表达式，等等. 第 6 节讨论随机徘徊取极限化为扩散过程的概要（扩散方程的形式解是随机徘徊的极限分布）.

在第 7 节中又回到初等的讨论，其内容是二维或多维随机徘徊，在这一节中将遇到新的现象. 第 8 节讨论一种全然不同的推

1）反之，有些新结果也能用第三章的方法来证明. 输光问题的利用无限多次反射的解法见问题 7—9.

广，即一种一维随机徘徊，其中质点不再限于每次移动一个单位步，而是可以跳跃地改变其位置，因此每次可以移动单位长度的任意倍. 这种广义的随机徘徊与瓦尔德的序贯抽样之间的关系已受到普遍的注意。

最后，必须强调指出，每个随机徘徊都表示一个特殊的马尔科夫链，故本章可部分地作为下章的一个导引，有几个随机徘徊的问题(例如弹性壁)将在下章重新论述.

最后一节(问题)中包含有正文部分的必不可少的补充以及其它方法的概要. 我们希望，所用的各种方法的比较将表明是极有教益的. （读者在查阅第三章的正文和图形时要设想时间轴是水平的，x 轴则与之垂直.）

2. 古典输光问题

在本节中我们将考虑本章开头所陈述的问题. 设 q_z 为赌徒最终[1]输光的概率，p_z 为他获胜的概率. 用随机徘徊的术语来说，q_z 与 p_z 分别为从 z 出发的质点将在 0 处与 a 处被吸收的概率. 我们将证明 $p_z + q_z = 1$，故我们不必考虑赌博永不终止的可能性.

在第一次试验后，赌徒的资本为 $z - 1$ 或 $z + 1$，故如果 $1 < z < a - 1$，则我们必定有

$$q_z = pq_{z+1} + qq_{z-1}. \tag{2.1}$$

如果 $z = 1$，则第一次试验就有可能输光，故此时须用 $q_1 = pq_2 + q$ 来代替 (2.1). 与此类似，如果 $z = a - 1$，则第一次试验就有可能获胜，故 $q_{a-1} = qq_{a-2}$. 为把这些方程统一起来，我们定义

$$q_0 = 1, \quad q_a = 0. \tag{2.2}$$

1) 严格地说，输光的概率是在无限延长的赌博的样本空间定义的，但我们能在 n 次试验的样本空间中来讨论. 在 n 次试验之前输光的概率随着 n 增加，因而其极限存在. 我们称此极限为“输光概率”. 本章中的所有概率都可以用这种方法来解释，而不必参照无限样本空间(参看第八章引言).

有了这个约定后，对于一切 $z = 1, 2, \cdots, a-1$，输光概率都满足(2.1)。

方程(2.1)是一差分方程，(2.2)则表示关于 q_z 的边界条件。我们将用特解法导出 q_z 的明显表达式，这种方法也可在更一般的情况使用。

首先假定 $p \neq q$. 容易证明，差分方程(2.1)有两个特解 $q_z = 1$ 及 $q_z = (q/p)^z$. 由此可知，对于任意常数 A 与 B，序列

$$q_z = A + B\left(\frac{q}{p}\right)^z$$

表示(2.1)的形式解. 边界条件(2.2)成立的充要条件是 A, B 满足两个线性方程 $A + B = 1$ 及 $A + B(q/p)^a = 0$. 于是

$$q_z = \frac{(q/p)^a - (q/p)^z}{(q/p)^a - 1} \tag{2.4}$$

是差分方程(2.1)满足边界条件(2.2)的一个形式解. 要证明(2.4)确是所求的输光概率，只需证明解的唯一性，即(2.1)的所有解都具有(2.3)的形式. 易知，对于已给(2.1)的一个任意解，可以选择两个常数 A 与 B，使得当 $z = 0$ 及 $z = 1$ 时，(2.3)与这个解一致. 从这两个解出发，依次以 $z = 1, 2, 3, \cdots$ 代入(2.1)即可求得所有其它值. 故如果两个解当 $z = 0$ 与 $z = 1$ 时一致，则它们必定恒等，因而每个解都具有(2.3)的形式.

当 $p = q = 1/2$ 时，以上的论证不再成立，这是因为，在这种情况下，两个形式特解 $q_z = 1$ 及 $q_z = (q/p)^z$ 恒等，故此时(2.4)没有意义. 然而，当 $p = q = 1/2$ 时，我们有第二个解 $q_z = z$. 故 $q_z = A + Bz$ 是(2.1)的依赖于两个常数的解. 为了满足边界条件，我们必须令 $A = 1$ 及 $A + Ba = 0$. 因此

$$q_z = 1 - \frac{z}{a}. \tag{2.5}$$

(形式上，在(2.4)中令 $p \to 1/2$，并用洛必大法则求极限，也可得出同一数值.)

于是我们证明了如下结论：所求的赌徒的输光概率，当 $p \neq q$

时由(2.4)给出，当 $p=q=1/2$ 时由(2.5)给出. 赌徒获胜的概率等于其对手的输光概率，故可由分别用 q, p, $a-z$ 来代替以上公式中的 p, q, a 而得到. 如前所述，容易看出，$p_z+q_z=1$.

我们将所得结果改述如下：设赌徒甲的初始赌本为 z，其对手则具有无限的赌本，且赌徒甲有权随意停止赌博，而其对手则愿意一直赌下去. 赌徒甲采取如下策略：当他输光全部赌本或使赌本增至 a 时(净赢 $a-z$)就停止赌博. 则 q_z 是他输光的概率，而 $1-q_z$ 则为他获胜的概率.

在采取这种策略的情况下，赌徒的最终赢利或损失是分别以概率 $1-q_z$ 与 q_z 取值 $a-z$ 与 $-z$ 的一个随机变量 $\mathbf{G}$. 期望赢利为

$$\mathbf{E}(\mathbf{G})=a(1-q_z)-z. \tag{2.6}$$

显然，当且仅当 $p=q$ 时，$\mathbf{E}(\mathbf{G})=0$. 这意味着，在采用上述策略时，“公平”赌博仍然是公平的，而且也不能将“不公平的”赌博改变为“公平”赌博.

由(2.5)可以看到，在 $p=q$ 的情况，一个初始赌本为 $z=999$ 的赌徒，在他输光赌本以前赢得一元的概率为 999/1000. 如果 $q=0.6$, $p=0.4$，则赌博确是不公平的，但由(2.4)[1]可知，在输光赌本以前赢得一元的概率仍然约为 2/3. 一般地说，如果赌徒的初始赌本很大(相对而言)，则在他输光以前赢得少量金额 $a-z$ 的可能性很大[2].

现在我们来研究改变赌注的影响. 将单位由一元改为半元是与将初始赌本加倍等价的. 对应的输光概率 q_z^* 可由在(2.4)中用 $2z$ 代替 z，用 $2a$ 代替 a 而得

---

1) 原书直接称(2.4)为输光赌本以前赢得一元的概率是不对的 ——译者注.

2) 某人过去每年都去蒙特卡罗(摩纳哥首都，以赌博著名 ——译者注.)赌博，而且总是赢回他假期的花费. 因此他坚信必有一种超机遇的魔力存在. 其实他的经验并不值得奇怪. 假定他所带的赌本是最终赢利的十倍，则每年他赌赢的机会约为 9/10 (根据(2.5)，这个概率的精确值为 10/11. 因此连赢十一次的概率亦为 $(1-1/11)^{11}\approx e^{-1}$ ——译者注.) 于是接连赌赢十次的概率为 $(1-1/10)^{10}\approx e^{-1}\approx 0.37$. 因此接连赌赢决不是不可能的.

$$q_z^* = \frac{(q/p)^{2a} - (q/p)^{2z}}{(q/p)^{2a} - 1} = q_z \cdot \frac{(q/p)^a + (q/p)^z}{(q/p)^a + 1}. \quad (2.7)$$

当 $q > p$ 时上式右端的分式大于 1，故 $q_z^* > q_z$. 我们将这个结论改述如下：如果赌注加倍而初始赌本不变，则输光概率对于成功概率 $p < 1/2$ 的赌徒来说是减少的，而对于他的对手(赌博对他有利)来说则是增加的. 例如，设赌徒甲有赌本 90 元，其对手乙有 10 元，而 $p = 0.45$，因而赌博对甲不利. 如果每次试验的赌注为一元，则表 1 表明，甲输光的概率约为 0.866. 在同一赌博中，如果

表 1　输光问题的举例说明

p	q	z	a	概率		期望值	
				输光	获胜	赢利	持续时间
0.5	0.5	9	10	0.1	0.9	0	9
0.5	0.5	90	100	0.1	0.9	0	900
0.5	0.5	900	1000	0.1	0.9	0	90000
0.5	0.5	950	1000	0.05	0.95	0	47500
0.5	0.5	8000	10000	0.2	0.8	0	16000000
0.45	0.55	9	10	0.210	0.790	−1.1	11
0.45	0.55	90	100	0.866	0.134	−76.6	765.6
0.45	0.55	99	100	0.182	0.818	−17.2	171.8
0.4	0.6	90	100	0.983	0.017	−88.3	441.3
0.4	0.6	99	100	0.333	0.667	−32.3	161.7

初始赌本为 z. 当赌本输光(损失 z 元)或增至 a (赢得 $a-z$ 元)时赌博结束.

赌注改为 10 元，则甲输光的概率下降到原概率的 1/4 以下，即约为 0.210. 可见，增加赌注的影响比我们的预料更为显著. 一般地说，如果每次试验的赌注为 k 元，则在(2.4)中用 z/k 代替 z，用 a/k 代替 a 即可求得输光概率；这个概率随着 k 的增加而减少. 故在赌注不变的赌博中，赌徒可以适当选择与他打算赢得事先确定的金额的目标相一致的大赌注，以使其输光概率减到最小. 这个结论的实际有效性，通常受到那些坚持认为每种"不公平的"赌博都是不合理的人的责难. 如果情况真是如他们所认为的那样，保

险事业就不可能存在，因为参加保险的谨慎司机显然在进行一场“不公平的”赌博．实际上，概率论中并没有什么定理会阻止这样一个司机参加保险．

3. 赌博持续时间的期望值

赌博持续时间的概率分布将在下节中导出．不过，它的期望值可用简单得多的方法导出，这种方法具有广泛的应用，故我们现在对它加以说明，而不惜稍有重复．

我们仍然考虑本章开头所表述的问题．我们假定赌博持续时间具有有限期望值这一事实是已知的．其严格证明将在下节中给出．

如果第一次试验的结果为成功，则赌博就像初始位置在 $z+1$ 处那样继续下去．故在第一次试验出现成功的假定下持续时间的条件期望为 $\mathbf{D}_{z+1}+1$．这表明，期望持续时间 $\mathbf{D}_z$ 满足如下差分方程

$$\mathbf{D}_z = p\mathbf{D}_{z+1} + q\mathbf{D}_{z-1} + 1, \quad 0 < z < a, \tag{3.1}$$

其边界条件为

$$\mathbf{D}_0 = 0, \quad \mathbf{D}_a = 0. \tag{3.2}$$

由于出现有常数项 1，故差分方程 (3.1) 是非齐次的．如果 $p \neq q$，则 $\mathbf{D}_z = z/(q-p)$ 为 (3.1) 的形式解．(3.1) 的任何两个解之差满足齐次方程 $\Delta_z = p\Delta_{z+1} + q\Delta_{z-1}$，又我们已知这个方程的所有解具有 $A + B(q/p)^z$ 的形式．由此可知，当 $p \neq q$ 时，(3.1) 的所有解具有如下形式

$$\mathbf{D}_z = \frac{z}{q-p} + A + B\left(\frac{q}{p}\right)^z. \tag{3.3}$$

边界条件 (3.2) 要求 $A + B = 0$ 及 $A + B(q/p)^a = -a/(q-p)$．解出 A，B，我们得

$$\mathbf{D}_z = \frac{z}{q-p} - \frac{a}{q-p} \cdot \frac{1-(q/p)^z}{1-(q/p)^a}. \tag{3.4}$$

当 $p=q=1/2$ 时，这个方法也不适用. 在这种情况，我们必须用 (3.1) 的一个解 $-z^2$ 来代替 $z/(q-p)$. 由此可知，当 $p=q=1/2$ 时，(3.1) 的所有解都具有 $\mathbf{D}_z=-z^2+A+Bz$ 的形式. 所需的满足边界条件 (3.2) 的解为

$$\mathbf{D}_z=z(a-z). \tag{3.5}$$

在古典输光问题中赌博持续时间的期望值，按 $p\neq q$ 或 $p=q=1/2$，分别由 (3.4) 或 (3.5) 给出.

应当指出，这个持续时间比我们所料想的要长得多. 设两个各有赌本 500 元的赌徒进行掷钱币的赌博，直到有一个输光为止，则平均持续时间为 250000 次试验. 如果一个赌徒仅有赌本一元，而其对手有赌本 1000 元，则平均持续时间为 1000 次试验. 进一步的例子可在表 1 中找到.

当 $a\to\infty$ 时的极限. 如果在输光概率的表达式中令 $a\to\infty$，则我们得到

$$q_z\longrightarrow\begin{cases}1, & \text{如果 } q\geqslant p,\\ (q/p)^z, & \text{如果 } q<p.\end{cases} \tag{3.6}$$

没有人会犹豫把这些极限解释为在和具有无限赌本的对手进行赌博时的输光概率，但在公理方面，半无限区间 $(0,\infty)$ 上的随机徘徊应当从概念本身来考虑. 在这种随机徘徊中，从 $z>0$ 出发的一个质点到达原点，与在自由随机徘徊中一个质点到达距其出发点 z 单位的左侧一位置的意义是一样的，故其概率相等. 这个概率我们已在第十一章第 3 节中算出，它与 (3.6) 是一致的：在和具有无限赌本的对手进行赌博时，输光的概率，当 $q>p$ 时为 1，当 $q<p$ 时为 $(q/p)^z$. 在第二种情况，谈论赌博持续时间的期望值是没有意义的，因为赌博可以永远继续下去. 当 $q>p$ 时赌博持续时间的期望值为极限 $z(q-p)^{-1}$，当 $q=p$ 时这个极限是无限的. 这与我们以前得到的关于对称随机徘徊的所有初过时间都具有无限期望值的结果是一致的. (这些结果的独立推导将在下节中给出.)

*4. 赌博持续时间及初过时间的母函数

我们将用母函数的方法来研究古典输光问题（即在 0 与 a 处

---

* 本节及与此有关的第 5 节在初次阅读时可以略去.

具有吸收壁的受约束随机徘徊）中的赌博持续时间．设初始位置为 $z\ (0 < z < a)$．令 $u_{z,n}$ 表示过程于第 n 步在吸收壁 0 处结束（即赌徒输光）的概率．第一步后的位置为 $z+1$ 或 $z-1$，故当 $1 < z < a-1$ 及 $n \geqslant 1$ 时我们有

$$u_{z,n+1} = pu_{z+1,n} + qu_{z-1,n}. \tag{4.1}$$

这是类似于 (2.1) 的差分方程，但它依赖于两个变数 z 与 n．和第 2 节类似，我们要定义边界值 $u_{0,n}$，$u_{a,n}$ 及 $u_{z,0}$，使得当 $z=1$，$z=a-1$ 及 $n=0$ 时 (4.1) 也成立．为此我们令

$$u_{0,n} = u_{a,n} = 0 \quad (n \geqslant 1) \tag{4.2}$$

及

$$u_{0,0} = 1, \quad u_{z,0} = 0 \quad (z > 0). \tag{4.3}$$

于是 (4.1) 对于所有满足 $0 < z < a$ 的 z 及所有的 $n \geqslant 0$ 成立．

现在我们引入母函数

$$U_z(s) = \sum_{n=0}^{\infty} u_{z,n}s^n. \tag{4.4}$$

以 s^{n+1} 乘上式两边并对 $n = 0, 1, 2, \cdots$ 求和，我们得

$$U_z(s) = psU_{z+1}(s) + qsU_{z-1}(s), \quad 0 < z < a, \tag{4.5}$$

又由方程 (4.2) 与 (4.3) 可得出边界条件

$$U_0(s) = 1, \quad U_a(s) = 0. \tag{4.6}$$

方程 (4.5) 是类似于 (2.1) 的差分方程，边界条件 (4.6) 则对应于 (2.2)．现在出现的新情况是，方程中的系数与未知函数 $U_z(s)$ 都依赖于 s，但就差分方程而论，s 仅仅是一个任意常数．如果我们能求得 (4.5) 的两个特解，则可以再次应用第 2 节中的方法．自然我们会考虑是否有形如 $U_z = \lambda^z(s)$ 的两个解．以此式代入 (4.5) 可知，$\lambda(s)$ 必须满足二次方程

$$\lambda(s) = ps\lambda^2(s) + qs, \tag{4.7}$$

其根为

$$\lambda_1(s) = \frac{1 + (1 - 4pqs^2)^{\frac{1}{2}}}{2ps},$$

$$\lambda_2(s)=\frac{1-(1-4pqs^2)^{\frac{1}{2}}}{2ps}. \tag{4.8}$$

(其中 $0<s<1$，且平方根取正值).

我们已经求出 (4.5) 的两个特解，于是可以断定(像第 2 节中那样)，对于任意的函数 $A(s)$ 与 $B(s)$，

$$U_z(s)=A(s)\lambda_1^z(s)+B(s)\lambda_2^z(s) \tag{4.9}$$

是 (4.5) 的一个解. 为了使 (4.9) 满足边界条件 (4.6)，我们必须有 $A(s)+B(s)=1$ 及 $A(s)\lambda_1^a(s)+B(s)\lambda_2^a(s)=0$，或

$$U_z(s)=\frac{\lambda_1^a(s)\lambda_2^z(s)-\lambda_1^z(s)\lambda_2^a(s)}{\lambda_1^a(s)-\lambda_2^a(s)}. \tag{4.10}$$

利用显然的关系 $\lambda_1(s)\lambda_2(s)=q/p$，上式可以简化为

$$U_z(s)=\left(\frac{q}{p}\right)^z\frac{\lambda_1^{a-z}(s)-\lambda_2^{a-z}(s)}{\lambda_1^a(s)-\lambda_2^a(s)}. \tag{4.11}$$

这就是所求的在第 n 次试验输光(在 0 处被吸收)的概率的母函数. 分别用 q, p 与 $a-z$ 代替 p, q, z 便得到对应的在 a 处被吸收的概率的母函数. 显然，赌博持续时间的母函数是这两个母函数之和.

$a=\infty$ 的情况

我们的方法对于 $a=\infty$ 的情况同样适用，这种情况对应于仅在原点处有一吸收壁的 $(0,\infty)$ 上的随机徘徊(或与具有无限赌本的对手进行赌博). 现在我们仅有一个边界条件 $U_0(s)=1$. (4.5) 的所有解都具有 (4.9) 的形式，但因为当 $0<s<1$ 时 $\lambda_1(s)>1$ 而 $\lambda_2(s)<1$，故只有当 $A(s)=0$ 时，$U_z(s)$ 才能有界. 于是所求的解为

$$V_z(s)=\lambda_2^z(s). \tag{4.12}$$

这是从 $z>0$ 出发的质点恰于第 n 次试验在原点被吸收的概率的母函数.

换句话说，在自由随机徘徊中，(4.12) 是初过位于初始位置左侧且与之相距 z 单位的点的时间的分布的母函数. 为了得出初过右侧的点的公式，只需交换 p 与 q. 由(4.8) 可知，在由原点出发

的自由随机徘徊中，初过点 $z>0$ 的时间的母函数为

$$\lambda^{z}(s)=\left(\frac{p}{q}\lambda_{2}(s)\right)^{z}=\lambda_{1}^{-z}(s). \tag{4.13}$$

对于特殊值 $z=1$，$\lambda(s)$ 是初过右侧一单位的母函数．由 0 出发初过任意的 $z>1$ 是由 0 到 1，由 1 到 2，…，由 $z-1$ 到 z 等初过时间的和，因而是具有同一母函数 $\lambda(s)$ 的 z 个独立随机变量之和．这正是（4.13）中的式子是一个母函数的 z 次幂的原因．

以 $s=1$ 代入（4.13）即可求得在对手具有无限赌本的情况的输光概率．这个概率当 $q\leqslant p$ 时为 1，当 $q\geqslant p$ 时为 $(q/p)^{z}$．

5. 明显表达式

现在我们用展开 $U_z(s)$ 为部分分式的方法来推导 $u_{z,n}$ 的明显公式．形式上，$U_z(s)$ 的表达式（4.11）依赖于一个平方根，但实际上 $U_z(s)$ 是一个有限函数．事实上，按二项式定理将（4.8）展开，我们看到，差 $\lambda_1^k(s)-\lambda_2^k(s)$ 是 s 的有理函数与 $(1-4pqs^2)^{\frac{1}{2}}$ 之积；这个根式作为一个因子同在（4.11）的分子与分母中出现，故 $U_z(s)$ 是两个多项式之比．分母的次数，当 a 为奇数时等于 $a-1$，当 a 为偶数时等于 $a-2$；分子的次数，当 $a-z$ 为奇数时等于 $a-1$，当 $a-z$ 为偶数时等于 $a-2$．在任何情况，分子次数与分母次数之差不可能大于 1．故如果分母没有重根，则 $u_{z,n}$ 可由第十一章的方程（4.8）来计算．

我们可以直接计算分母的根及对应的系数，但如果通过

$$\frac{1}{\cos\phi}=2(pq)^{\frac{1}{2}}s \tag{5.1}$$

引进一个新独立变量 ϕ，则运算可以简化．由（4.8）我们有

$$\lambda_{1,2}(s)=\left(\frac{q}{p}\right)^{\frac{1}{2}}(\cos\phi\pm i\sin\phi)=\left(\frac{q}{p}\right)^{\frac{1}{2}}e^{\pm i\phi}, \tag{5.2}$$

故由（4.1）得

$$U_z(s)=\left(\frac{q}{p}\right)^{\frac{1}{2}z}\frac{\sin(a-z)\phi}{\sin a\phi}. \tag{5.3}$$

分母的根显然为 $\phi = 0, \pi/a, 2\pi/a, \cdots$. s 的对应值为

$$s_\nu = \frac{1}{2(pq)^{\frac{1}{2}} \cos \nu\pi/a}. \tag{5.4}$$

令 $\nu = 0, 1, \cdots, a$, 即得 s_ν 的所有可能值. 不过, 对应于 $\nu = 0$ 与 $\nu = a$ 的值 $\phi = 0, \pi$, 也是(5.3)中分子的根, 又如果 a 为偶数, 则当 $\nu = \frac{1}{2}a$ 时,不存在与之对应的数 s_ν. 因此,当 a 为奇数时, 令 $\nu = 1, 2, \cdots, a-1$, 即得所有 $a-1$ 个根 s_ν; 当 a 为偶数时, $\nu = \frac{1}{2}a$ 的值必须除外. 我们不必考虑也是分子的根的那些 s_ν, 对它们来说,由(5.6)显然有 $\rho_\nu = 0$.

我们知道

$$\left(\frac{q}{p}\right)^{\frac{1}{2}z} \frac{\sin (a-z)\phi}{\sin a\phi}$$

$$= As + B + \frac{\rho_1}{s_1 - s} + \cdots + \frac{\rho_{a-1}}{s_{a-1} - s}. \tag{5.5}$$

为求 ρ_ν, 以 $s_\nu - s$ 乘两边并令 $s \to s_\nu$. 像在第十一章(4.5)中一样,我们有(令 $\phi_\nu = \pi\nu/a$)

$$\rho_\nu = -\left(\frac{q}{p}\right)^{\frac{1}{2}z} \frac{\sin (a-z)\pi\nu/a}{a \cdot \cos \nu\pi \cdot (d\phi/ds)_{s=s_\nu}}$$

$$= \left(\frac{q}{p}\right)^{\frac{1}{2}z} \frac{\sin z\pi\nu/a \cdot \sin \pi\nu/a}{2a(pq)^{\frac{1}{2}} \cos^2 \pi\nu/a}. \tag{5.6}$$

所以当 $n > 1$ 时,由(5.5)可得 s^n 的系数 $u_{z,n}$ 的表达式

$$u_{z,n} = a^{-1} 2^n p^{\frac{1}{2}(n-z)} q^{\frac{1}{2}(n+z)} \sum_{\nu=1}^{a-1} \cos^{n-1} \frac{\pi\nu}{a} \cdot \sin \frac{\pi\nu}{a} \cdot \sin \frac{\pi z\nu}{a}. \tag{5.7}$$

(严格地说, 当 a 为偶数时, $\nu = \frac{1}{2}a$ 这一项应于略去, 但它为 0 故写上也没关系.)

当 $n > 1$ 时, 公式(5.7)表示在第 n 次试验输光(吸收)的概率. 这个公式最先是由拉格朗日得到的, 现在已有很多不同的推

导方法[1]. 尽管它有光荣的历史且有选入教科书的价值,但这个公式还是常被重复发现. 另一个明显表达式见问题 13. 极限形式见第 6 节及问题 14(对于反射壁的类似公式在第十六章第 3 节中导出).

如果令 $a \to \infty$, 则 (5.7) 中的和可以解释为逼近一个积分的黎曼和. 用这个方法可以求得, 在与一个具有无限赌本的对手进行赌博时(在 0 处具有单一吸收壁), 具有初始赌本 $z > 0$ 的赌徒恰好在第 n 步输光的概率 $w_{z,n}$ 为

$$w_{z,n} = 2^n p^{\frac{1}{2}(n-z)} q^{\frac{1}{2}(n+z)} \int_0^1 \cos^{n-1}\pi x \sin \pi x \sin \pi x z \cdot dx. \tag{5.8}$$

这个积分可用初等方法[2]表示如下

$$w_{z,n} = \frac{z}{n} \binom{n}{\frac{1}{2}(n-z)} p^{\frac{1}{2}(n-z)} q^{\frac{1}{2}(n+z)}; \tag{5.9}$$

其中如果 $\frac{1}{2}(n-z)$ 不是区间 $[0, n]$ 中的整数,则二项式系数解释为 0. 可以求得对应的母函数为 $\lambda_2^z(s)$(见第 4 节末).

6. 趋向极限; 扩散过程

我们曾指出, 随机徘徊的模型可以作为扩散理论与布朗运动的粗略近似, 其中细小质点受到大量分子的冲击. 每次冲击的影响是微不足道的,但很多微小作用的迭加则产生出可观测的运动. 因此, 我们现在要研究一种随机徘徊, 在其中每一个别步都极微小,但它们迅速地接连出现. 在取极限时,这种过程就像是一种连续运动. 有趣的是, 在取极限时我们的公式仍有意义且与扩散理论中具有重要物理意义的公式一致, 后者可以在普遍得多的条件

1) 艾利斯 (R. E. Ellis) 在 [22] 中用三角插值法给出了一种初等推导.

2) 当 $p = q = 1/2$ 时,公式 (5.9) 化归为第三章中的初过时间分布公式 (4.11). 证明 (5.8) 与 (5.9) 相一致决不是容易的. 也许最简单的方法是证明这两个公式都表示差分方程 (4.1) 满足在原点处的边界条件 (4.2)—(4.3) 的解.

下用更加简捷的方法导出.[1] 这部分地说明，为什么随机徘徊的模型虽然粗糙，描述扩散过程却相当成功；仅仅极限情况具有物理意义，但各种离散模型都导致同一极限公式. 这里的情形在很多方面类似于中心极限定理的条件，其中很多偶然成分的累积效应实际上是与各个成分的性质无关的.

我们首先考虑从原点出发的自由随机徘徊，并令 $v_{x,n}$ 为质点在第 n 步位于 x 处的概率. 如果在 n 步中有 r 步向右，$n-r$ 步向左，则总位移为 $r-(n-r)=2r-n$ 单位. 仅当 n 与 x 同为偶数或同为奇数时，位移才能等于 x（这意味着偶数步后坐标 x 是一偶数）. 由 n 步中选取 r 步共有 $\binom{n}{r}$ 种方法，故

$$v_{x,n}=\binom{n}{\frac{1}{2}(n+x)}p^{\frac{1}{2}(n+x)}q^{\frac{1}{2}(n-x)}; \tag{6.1}$$

其中如果 $\frac{1}{2}(n+x)$ 不是区间 $[0,n]$ 中的整数，则二项式系数解释为 0.

利用导出差分方程 (4.1) 及边界条件 (4.2) 与 (4.3) 的方法可以给出 (6.1) 的另一种推导. 能够证明，$v_{x,n}$ 必须满足差分方程

$$v_{x,n+1}=pv_{x-1,n}+qv_{x+1,n} \tag{6.2}$$

及边界条件

$$v_{0,0}=1,\quad v_{x,0}=0\quad(x\neq 0). \tag{6.3}$$

设已给 (6.3)，在 (6.2) 中依次令 $n=1,2,\cdots$，则首先得到所有的值 $v_{x,1}$，然后又依次可得 $v_{x,2}$，$v_{x,3}$，$\cdots$. 这表明条件 (6.2) 与 (6.3) 唯一地确定 $v_{x,n}$. 另一方面，容易看出，(6.1) 是一个解.

现在我们改变长度单位，使得每步长为 Δx，又假定任何接连

---

1) 本节中的极限公式与现在的古典爱因斯坦-维纳 (Einstein-Wiener) 理论中的公式一致. 此处没有考虑新的、更为严密的理论[乌伦贝克 (Uhlenbeck)、奥恩斯坦 (Ornstein)]. 发现随机徘徊与扩散之间的联系主要归功于巴切利尔 (L. Bachelier). 他的工作往往是启发性的，但他也导出很多新结果. 柯尔莫果洛夫的马尔科夫型随机过程的理论在很大程度上是以巴切利尔的思想为基础的. 特别见 [23].

两步之间的时间为 Δt. 在时间 t 内，质点大约移动了 $t/\Delta t$ 步，又现在位移 x 相当于 $x/\Delta x$ 个单位. 仅仅 Δx 与 Δt 的倍数才表示有意义的坐标，但如果 $\Delta x \to 0$, $\Delta t \to 0$, 则每个位移和所有的时间都是可能的.

如果 Δx 与 Δt 按任意的方式趋向于 0，则不可能得到可察觉的结果，因为在时间 t 中最大可能位移等于 $t\Delta x/\Delta t$，所以如果 $\Delta x/\Delta t \to 0$，则在取极限时不可能有运动. 用物理的话来说，我们必须使 x 和 t 的尺度保持适当的比例，不然的话，过程将退化，而方差则趋向于零或无穷. 为了求得适当的比值，我们注意，在时间 t 内的总位移是大约 $t/\Delta t$ 个相互独立的随机变量之和，其中每个具有均值 $(p-q)\Delta x$ 及方差 $\{1-(p-q)^2\}(\Delta x)^2 = 4pq(\Delta x)^2$. 于是在时间 t 的总位移的均值与方差分别约为 $t(p-q)\Delta x/\Delta t$ 及 $4pqt(\Delta x)^2/\Delta t$. 为了得到合理的结果，我们必须令 Δx 与 Δt 按这样一种方式趋向于 0，使得对于所有的 t，上述均值与方差保持有限. 方差的有限性要求 $(\Delta x)^2/\Delta t$ 保持有界，均值的有限性则要求 $p-q$ 的大小与 Δx 同阶. 因此我们令

$$\frac{(\Delta x)^2}{\Delta t} = 2D, \quad p = \frac{1}{2} + \frac{c}{2D}\Delta x, \quad q = \frac{1}{2} - \frac{c}{2D}\Delta x, \tag{6.4}$$

其中 D 与 c 为常数. D 的值仅仅引入一个尺度因子；为数学上简单起见，最好是令 $D = 1$，但为了易于和物理理论比较，我们不指定 D 的值. 常数 D 与 c 分别为扩散系数与趋向. 如果 $c = 0$，则随机徘徊是对称的；一般的说，c 的符号确定趋向的方向. 在取极限时，p 与 q 趋向于 1/2；在任何其它规定条件下，质点都会如此迅速地趋向远处，以至位移有限的概率趋向于零.

在规定条件 (6.4) 的限制下，我们令 $\Delta x \to 0$, $\Delta t \to 0$. 由时刻 $t \approx n\Delta t$ 的总位移由 n 次伯努利试验决定，故 $v_{x,n}$ 的极限形式由正态分布给出. 对于固定的 Δx，这位移是有限多个独立随机变量之和，其均值为 $t(p-q)\Delta x/\Delta t = 2ct$；其方差为 $4pqt(\Delta x)^2/\Delta t \to 2Dt$. 所以在时刻 t 位移位于 x_0 与 $x_1(x_0 < x_1)$ 之间的概率趋向于

$$(2\pi)^{-\frac{1}{2}}\int_{y_0}^{y_1} e^{-\frac{1}{2}\lambda^2}\,d\lambda, \tag{6.5}$$

其中 $y_1=(x_1-2ct)/(2Dt)^{\frac{1}{2}}$, $y_0=(x_0-2ct)/(2Dt)^{\frac{1}{2}}$.

利用通常的函数记号可将方程 (6.2) 写成如下形式

$$v(x,t+\Delta t)=p\cdot v(x-\Delta x,t)+q\cdot v(x+\Delta x,t).$$

按照泰勒定理展开到二阶项,我们形式地便有

$$\Delta t\cdot\frac{\partial v(x,t)}{\partial t}=(q-p)\Delta x\cdot\frac{\partial v(x,t)}{\partial x}+\frac{(\Delta x)^2}{2}\frac{\partial^2 v(x,t)}{\partial x^2}+\cdots. \tag{6.6}$$

利用 (6.4), 取极限得

$$\frac{\partial v(x,t)}{\partial t}=-2c\cdot\frac{\partial v(x,t)}{\partial x}+D\cdot\frac{\partial^2 v(x,t)}{\partial x^2}. \tag{6.7}$$

这就是福克-普朗克 (Fokker-Planck) 的具有趋向的扩散方程, 它可由更为一般的、更有说服力的假设导出. 在通常的理论中, 解 (6.5) 是由 (6.7) 导出的, 而我们是在同一极限过程中得到这两个公式的. 我们的程序只是启发性的, 但能严格地予以证明. 对于离散随机徘徊的所有公式都可以类似地考虑其极限.

作为一个进一步的例子, 考虑初过概率的极限形式. 为简单起见, 我们首先考虑对应于单一壁的公式 (5.9). 两个量 $w_{z,n}$ 与 $w_{z,n+1}$ 中必有一个为零. 和 $w_{z,n}+w_{z,n+1}$ 渐近地表示在时间间隔 $(t, t+2\Delta t)$ 中吸收的概率. 我们将证明, $w_{z,n}+w_{z,n+1}\sim f(z,t)(2\Delta t)$, 其中 $f(z,t)$ 为连续函数. 于是在任何时间间隔 (t_1, t_2) 中吸收的极限概率为 $f(z,t)$ 在此间隔上的积分. 当 $n-z$ 为偶数时,我们有 $w_{z,n+1}=0$, 为求 $f(z,t)$, 我们必须用 $z/\Delta x$ 代替(5.9)中的 z, 用 $t/\Delta t$ 代替 n, 并应用 (6.4). 利用二项分布的正态近似及方程 (6.9) 容易求得[1]

$$f(z,t)\sim\frac{z}{2(\pi Dt^3)^{\frac{1}{2}}}e^{-\frac{1}{4}(z+2ct)^2/Dt}. \tag{6.8}$$

---

1) 在 $c=0$ 的对称情况(即 $p=q$), 公式 (6.8) 与第三章 (8.c) 中用初等方法导出的初过时间的极限分布一致.

这就是(5.9)的极限形式；它又与扩散理论中的对应公式一致．事实上，容易证明，$f(-x, t)$ 是(6.7)的一个解．(在 $w_{z,n}$ 的定义中，变量 z 起着 $v_{x,n}$ 中的 $-x$ 的作用．)

类似的论证适用于(5.7)．容易看出，如果 $n-z$ 为奇数，则这个公式中相应于 $\nu=k$ 与 $\nu=a-k$ 的两项相消，如果 $n-z$ 为偶数，则这两项相加．故将(5.7)中展布在 $1 \leqslant \nu < a/2$ 上的和加倍，就可得到极限形式 $f(z, t) \sim (u_{z,n}+u_{z,n+1})/(2\Delta t)$．分别用 $z/\Delta x$, $a/\Delta x$, $t/\Delta t$ 代替 z, a, n，并注意，对于固定的 ν 有

$$\sin\frac{\pi\nu\Delta x}{a} \sim \frac{\pi\nu\Delta x}{a},$$

$$\left(\cos\frac{\pi\nu\Delta x}{a}\right)^{t/\Delta t} \sim \left(1-\frac{D\pi^2\nu^2\Delta t}{a^2}\right)^{z/\Delta t} \sim e^{-D\pi^2\nu^2 t/a^2}, \quad (6.9)$$

$$(4pq)^{t/2\Delta t}\left(\frac{q}{p}\right)^{z/2\Delta x} \sim e^{-c(ct+z)/D},$$

我们就形式地得到如下的极限形式

$$f(z, t) \sim 2\pi D a^{-2} e^{-c(ct+z)/D}\sum_{\nu=1}^{\infty}\nu e^{-D\pi^2\nu^2 t/a^2}\sin\frac{\pi z\nu}{a}. \quad (6.10)$$

以上形式地取极限之所以是正当的，是由于一致收敛性：当 ν 很大时，(6.10)及原来的和(5.7)中相应于 ν 的项可以忽略不计(其中我们有 $\nu < a/2$)．

在扩散理论中，(6.10)称为初过的弗思(Fürth)公式，它可直接由福克-普朗克方程导出．在自由扩散中，(6.10)在时间间隔 (t_1, t_2) 上的积分给出从 $z>0$ 出发的质点在先前未经过壁 a 的情况下在该时间间隔内第一次到达原点的概率．

*7. 平面和空间中的随机徘徊

在二维随机徘徊中，质点每次沿平行于 x 轴与 y 轴的四个方

---

* 本节讨论专门的论题，初读时可以略去．

向之一移动单位步．对于从原点出发的质点来说，其可能位置是平面上所有坐标为整数的点．每一位置有四个邻近位置．类似，在三维的情况，每一位置有六个邻近位置．为了定义随机徘徊，必须指定对应的四个或六个概率．为简单起见，我们仅考虑所有方向具有相同概率的对称情况．问题的复杂性要比一维的情况大得多，因为现在限制质点的区域可以具有任意形状，故复杂的边界取代了一维情况的单点壁．

我们首先讨论波利亚的一个有趣的定理．[1)]

定理．在对称一维与二维随机徘徊中，质点迟早将返回其初始位置的概率为 1（因而返回其初始位置无穷多次的概率亦为 1）．然而，在三维的情况，这个概率仅约为 0.35（于是返回次数的期望值为 $0.65\sum k(0.35)^k = 0.35/0.65 \approx 0.53$）．

在证明这个定理之前，我们先给出波利亚所作的两种说明．首先，由这个定理显然可以推出，在一维与二维的情况，质点经过每一可能位置无穷多次的概率为 1；然而，在三维的情况，上述结论却不成立．于是这个定理，在某种意义上，就二维的情况为"条条道路通罗马"这一谚语提供了根据．

另一方面，考虑独立地随机徘徊的两个质点，假定它们每次移动同时发生．它们是否迟早总会相遇？为叙述简便起见，我们定义两个可能位置的位移为从一个位置到另一个位置的最少步数．（于是距离＝对应坐标差的绝对值之和）．如果两个质点各移动一步，则它们之间的相互距离或者保持不变，或者改变两个单位，故它们的距离或者在所有时刻都是偶数，或者都是奇数．在第二种情况，这两个质点永远不可能占据同一位置．在第一种情况，容易看到，它们在第 n 步相遇的概率等于第一个质点在 $2n$ 步内到达第二个质点的初始位置的概率．因此由我们的定理可知，在二维的情况（但在三维的情况则不成立），两个质点必有无穷多次占据同

1) 见 [24]．数值 0.35 是由麦克雷（W. H. McCrea）与惠普尔（F. J. W. Whipple）算出的，见 [25]．

一位置．如果两质点的初始距离为奇数，则用类似的方法可以证明，它们将有无穷多次占据相邻的位置．如果这两种情况都称为相遇，则我们的定理断定：在一维和二维的情况，两个质点必定相遇无穷多次，但在三维的情况，两个质点永不相遇具有正的概率．

证．对于一维的情况，这个定理已在第十三章例(3.b)中证明过了，只是在那里我们没有提到对称随机徘徊，而是用掷钱币的例子来叙述的．二维和三维的情况可遵循同一路线来证明．设 u_n 为质点于第 n 次试验返回初始位置的概率．根据第十三章第3节定理2，我们必须证明在二维的情况 $\sum u_n$ 发散，而在三维的情况 $\sum u_n \approx 0.53$．在二维的情况，仅当沿 x 轴与 y 轴的正向移动的步数分别等于沿 x 轴与 y 轴的负向移动的步数时，质点才可能返回初始位置．故如果 n 为奇数，则 $u_n = 0$；又利用第六章的多项分布公式(9.2)有

$$u_{2n} = \frac{1}{4^{2n}} \sum_{k=0}^{2n} \frac{(2n)!}{k!k!(n-k)!(n-k)!}$$

$$= \frac{1}{4^{2n}} \binom{2n}{n} \sum_{k=0}^{n} \binom{n}{k}^2, \tag{7.1}$$

根据第二章公式(12.11)，上式右端的式子等于 $4^{-2n}\binom{2n}{n}^2$．由史特令公式可知，u_{2n} 的大小与 $1/n$ 同阶，故 $\sum u_{2n}$ 发散．

在三维的情况，用类似的方法可以求得

$$u_{2n} = \frac{1}{6^{2n}} \sum_{j,k} \frac{(2n)!}{j!j!k!k!(n-j-k)!(n-j-k)!}, \tag{7.2}$$

其中的和是对所有满足 $j+k \leqslant n$ 的 j, k 来求的．易证

$$u_{2n} = \frac{1}{2^{2n}} \binom{2n}{n} \sum_{j,k} \left\{ \frac{1}{3^n} \frac{n!}{j!k!(n-j-k)!} \right\}^2. \tag{7.3}$$

在大括号内的项表示三项分布，我们知道，其和为1．故其平方和小于该括号内的最大项，而后者在 j 与 k 最接近 $n/3$ 时达到．由史特令公式可知，这个最大项的大小与 n^{-1} 同阶，故 u_{2n} 的大小与

$n^{-\frac{3}{2}}$ 同阶，由此知 $\sum u_{2n}$ 收敛．证毕．

波利亚定理类似于与第十三章例 (3.c) 中所讨论的投掷多个钱币有关的一些事实．

最后我们讨论推广吸收壁概念的另一问题．考虑两维的情况，这时代替区间 $0 \leqslant x \leqslant a$ 的是一个平面区域 D，即坐标为整数的一个集合．每个点都有四个邻近位置，但对于 D 中的某些点，有一个或多个邻近位置在 D 之外．这种点构成 D 的边界，所有其它的点则称为内点．在一维情况，边界由两个壁构成，我们的问题在于求从 z 出发的质点将在到达 a 之前到达边界点 0 的概率．类似，现在我们要求质点将在到达某段边界之外的任何点之前到达该段边界的概率．这意味着我们将所有的边界点分成两个集合 B' 与 B''．设 (x, y) 是一个内点，我们要求从 (x, y) 出发的质点将在到达 B'' 中的点之前到达 B' 中的一个点的概率 $u(x,y)$．特别，如果 B' 由一个点组成，则 $u(x, y)$ 为质点迟早将在这一特殊点被吸收的概率．

设 (x, y) 为一内点．在第一步后，质点从 (x, y) 移动到四个邻近位置 $(x\pm1, y)$，$(x, y\pm1)$ 之一处，如果这四个邻近位置都是内点，则我们有

$$u(x, y) = \frac{1}{4}[u(x+1, y) + u(x-1, y) + u(x, y+1) + u(x, y-1)]. \tag{7.4}$$

这是一个偏差分方程，它相应于一维情况的方程 (2.1) (其中 $p = q = 1/2$)．如果 $(x+1, y)$ 是一边界点，则相应于它的项 $u(x+1, y)$ 必须按 $(x+1, y)$ 属于 B' 或属于 B'' 而分别用 1 或 0 来代替．因此，如果当边界点 (ξ, η) 在 B' 中时我们令 $u(\xi, \eta) = 1$，当 (ξ, η) 在 B'' 中时令 $u(\xi, \eta) = 0$，则 (7.4) 对所有的内点成立．这个约定相应于一维情况的边界条件 (2.2)．

(7.4) 是关于未知数 $u(x, y)$ 的一组线性方程；每个内点对应一个未知数及一个方程．这组线性方程是非齐次的，因为在其中至少出现 B' 中的一个边界点，而这个点使得方程的右边出现一项

1/4. 如果区域是有限的，则方程的个数与未知数的个数是一样的，故这方程组有唯一解的充分必要条件是对应的齐次方程组(对于所有边界点令 $u(\xi, \eta)=0$) 仅有恒等于 0 的解. 由于 $u(x, y)$ 是四个邻近值 $u(x\pm1, y)$, $u(x, y\pm1)$ 的平均，故它不可能比所有这四个值都大，也不可能比它们都小. 换句话说，$u(x, y)$ 既不可能是严格意义下的极大值，也不可能是严格意义下的极小值. 故最大值与最小值在边界上达到. 因此，如果所有的边界值为 0，则 $u(x, y)$ 在所有的内点亦为 0. 这就证明了 (7.4) 的解的存在唯一性. 因为边界值为 0 与 1，故 $u(x, y)$ 的所有值在 0 与 1 之间，这正是它们作为概率所应满足的条件. 根据关于无限马尔科夫链的一个一般定理，我们将看到，以上陈述对于无限区域的情况也成立[1].

8. 广义一维随机徘徊(序贯抽样)

现在再来讨论一维的情况，但放弃质点每次移动单位步的限制. 代替的是，我们假定，质点在每一步从 x 移动到 $x+k$ 的概率为 p_k，其中整数 k 可以为零、为正或为负. 我们研究如下的输光问题：质点从位置 z $(0<z<a)$ 出发；我们要寻求质点在它到达任何 $\geqslant a$ 的位置之前到达 $\leqslant 0$ 的某个位置的概率 u_z. 换句话说，质点在时刻 n 的位置是 x 轴上的点 $z+\mathbf{X}_1+\mathbf{X}_2+\cdots+\mathbf{X}_n$，其中 $\{\mathbf{X}_k\}$ 是具有共同分布 $\{p_\nu\}$ 的相互独立的随机变量；当 $\mathbf{X}_1+\cdots+\mathbf{X}_n\leqslant 0$ 或 $\mathbf{X}_1+\cdots+\mathbf{X}_n\geqslant a-z$ 第一次出现时过程停止.

这个问题，由于它与序贯抽样有关而受到普遍的注意. 在那里，$\mathbf{X}_k$ 表示样本或观察的某种特征. 测量一直进行到和 $\mathbf{X}_1+\cdots+\mathbf{X}_k$ 落在两个事先选定的界限(即 $-z$ 与 $a-z$) 之外为止. 第一种情况导致拒绝，第二种情况导致接受. 首次讨论这种抽样

1) 仅仅得到了几种特殊情况的明确解答，这些解法都极为复杂. 关于矩形区域、无限窄条区域等的解法见 [25].

的是巴特基（W. Bartky）[1]；瓦尔德给出了一般理论的概要，以上的说明也是他给出的[2]。

不失一般性，我们将假定，沿正向和负向移动都是可能的．否则 $u_z = 0$ 或 $u_z = 1$ 对所有的 z 成立．

第一步输光的概率显然为

$$r_z = p_{-z} + p_{-z-1} + p_{-z-2} + \cdots \tag{8.1}$$

（这个量可以为 0）．仅当质点移动到的位置 x 满足 $0 < x < a$ 时，随机徘徊才能继续下去；质点从 z 跳到 x 的概率为 p_{x-z}，其后输光的概率为 u_x．故有

$$u_z = \sum_{x=1}^{a-1} u_x p_{x-z} + r_z. \tag{8.2}$$

这里我们又有 $a-1$ 个线性方程，其中有 $a-1$ 个未知数 u_z．这组方程是非齐次的，因为至少当 $z = 1$ 时的概率 r_1 是异于 0 的（我们曾假定沿负向移动是可能的，由此显然有 $r_1 > 0$）．

我们要证明对应的齐次方程组

$$u_z = \sum_{x=1}^{a-1} u_x p_{x-z} \tag{8.3}$$

仅有恒等于 0 的解．

事实上，如果它有另一组解，则值 u_z 中必有一个绝对值最大者，设此值为 $u_z = M > 0$．首先假定 $p_{-1} \neq 0$．因为 (8.3) 中诸系数之和最多为 1，故仅当所有确实在右边出现的 u_x[3]（其系数异于 0）都等于 M 且其系数之和为 1 时，方程 (8.3) 才可能成立．因为 $p_{-1} \neq 0$，故 $u_{z-1} = M$．用同样的方法可依次证明 $u_{z-2} = u_{z-3} = \cdots = u_1 = M$．然而，当 $z = 1$ 时，(8.3) 中诸系数 p_{x-z} 之和小于 1，故 M 必须为 0．以上的论证显然也适用于 $p_{-1} = 0$ 的情况，因为我们可用某个正系数 $p_k (k < 0)$ 来代替 p_{-1}．

---

1) 见 [26]，并参见第十五章例 (2.j)．

2) 见 [27]．本书所叙述的方法与瓦尔德的方法不同．还可看瓦尔德的书 [28]．

3) 原书误为 p_{x-z} ——译者注．

由此推出，(8.2) 有唯一解，于是我们的问题得以解决．方程 (8.2) 起着差分方程 (2.1) 的作用．通过引进边界条件

$$u_x = 1, \quad \text{如果 } x \leqslant 0,$$
$$u_x = 0, \quad \text{如果 } x \geqslant a \tag{8.4}$$

可使书写简化．于是 (8.2) 可以写成如下形式

$$u_z = \sum u_x p_{x-z}, \tag{8.5}$$

其中和是对所有的 x 来求的（由 (8.4) 中的第二个条件可知，当 $x \geqslant a$ 时相应的项为0；由第一个条件可知，对应于 $x \leqslant 0$ 的各项之和为 r_z）．

当 a 很大时，直接解 $a-1$ 个线性方程是麻烦的，这时最好是利用与第 2 节中的程序类似的特解法．当概率分布 $\{p_k\}$ 中的正项较少时，这种方法是可行的．设仅当 $-\nu \leqslant k \leqslant \mu$ 时，p_k 异于 0，故沿正向和负向的最大的可能跳跃分别为 μ 与 ν．特征方程

$$\sum p_k s^k = 1 \tag{8.6}$$

等价于一个 $\nu+\mu$ 阶的代数方程．如果 s 为 (8.6) 的一个根，则对于所有 z，$u_z = s^z$ 是 (8.5) 的一个形式解，但这个解不满足边界条件 (8.4)．如果 (8.6) 有 $\mu+\nu$ 个不同的根 $s_1, s_2, \cdots$，则对于所有的 z，线性组合

$$u_z = \sum A_k s_k^z \tag{8.7}$$

也是 (8.5) 的一个解，但我们必须适当选择常数 A_k，使得 (8.7) 满足边界条件．现在对于 $0 < z < a$，仅有满足 $-\nu+1 \leqslant x \leqslant a+\mu-1$ 的 x 值在(8.5)中出现．故使边界条件(8.4)对于 $x = 0, -1, -2, \cdots, -\nu+1$ 及 $x = a, a+1, \cdots, a+\mu-1$ 成立就够了．因此我们共有 $\mu+\nu$ 个条件．如果 s_k 是 (8.5) 的二重根，则我们少了一个常数，但容易看出，在这种情况，$u_z = zs_k^z$ 是另一个形式解．在每种情况，$\mu+\nu$ 个边界条件确定 $\mu+\nu$ 个任意常数．

例．假定每移动一步，质点到达四个最近的位置之一，并设 $p_{-2} = p_{-1} = p_1 = p_2 = 1/4$．特征方程 (8.6) 为 $s^{-2} + s^{-1} + s + s^2 = 4$．为解此方程，我们令 $t = s + s^{-1}$；作这个变换后，方程变

为 $t^2+t=6$，其根为 $t=2, -3$. 由 $t=s+s^{-1}$ 中解出 s，我们得到四个根：

$$s_1=s_2=1,\quad s_3=\frac{-3+5^{\frac{1}{2}}}{2}=s_4^{-1},\quad s_4=\frac{-3-5^{\frac{1}{2}}}{2}=s_3^{-1}. \tag{8.8}$$

因为 s_1 是二重根，故在这种情况 (8.5) 的一般解为

$$u_z=A_1+A_2z+A_3s_3^z+A_4s_4^z. \tag{8.9}$$

边界条件为 $u_0=u_{-1}=1$ 及 $u_a=u_{a+1}=0$. 它们导出关于系数 A_j 的四个线性方程，最后得出的解为

$$u_z=1-\frac{z}{a}+\frac{(2z-a)(s_3^a-s_4^a)-a(s_3^{2z-a}-s_4^{2z-a})}{a\{(a+2)(s_3^a-s_4^a)-a(s_3^{a+2}-s_4^{a+2})\}}, \tag{8.10}$$

其中 s_3 与 s_4 由 (8.8) 给出.

数值近似. 通常要求出所有的根是麻烦的，但用极简单的方法就可得到相当满意的近似. 首先考虑概率分布 $\{p_k\}$ 的均值为零的情况. 这时特征方程 (8.6) 有二重根 $s=1$，又 $A+Bz$ 为 (8.5) 的一个形式解. 当然要满足 $\mu+\nu$ 个边界条件 (8.4)，仅有两个常数 A 与 B 是不够的. 不过，如果我们确定 A 与 B，使得 $A+Bz$ 当 $z=a+\mu-1$ 时为 0，当 $z=0$ 时为 1，则当 $x\leqslant 0$ 时 $A+Bx\geqslant 1$，当 $a\leqslant x<a+\mu$ 时 $A+Bx\geqslant 0$，故如果用"大于或等于"代替 (8.4) 中的等号，则 $A+Bz$ 满足这个边界条件. 所以差 $A+Bz-u_z$ 是 (8.5) 的具有非负边界值的形式解，故有 $A+Bz-u_z\geqslant 0$. 与此类似，确定 A 与 B，使得 $A+Bz$ 当 $z=a$ 时为 0，当 $z=-\nu+1$ 时为 1，则可得出 u_z 的下界. 因此我们有

$$\frac{a-z}{a+\nu-1}\leqslant u_z\leqslant\frac{a+\mu-z-1}{a+\mu-1}. \tag{8.11}$$

如果 a 很大(相对 $\mu+\nu$ 而言)，则这个估值是极好的. (当然，$u_z\approx(1-z/a)$ 是一个较好的近似公式，但它未给出误差限.)

其次，考虑分布 $\{p_k\}$ 的均值不为零的一般情况. 这时特征方程 (8.6) 有单根 $s=1$. 当 $s\to 0$ 及 $s\to\infty$ 时，(8.6) 的左边趋向于 ∞. 当 s 为正时，$y=\sum p_k s^k$ 是连续凸曲线，又因为它在 $s=1$ 处与直线 $y=1$ 相交，故必定还有一个交点. 因此特征方程 (8.6) 恰有两个正根 1 与 s_1. 和前面一样，$A+Bs_1^z$ 是 (8.5) 的一个形式解，又我们能将前面的论证应用于这个解. 于是在这种情况，我们求得

$$\frac{s_1^a - s_1^z}{s_1^a - s_1^{-\nu+1}} \leqslant u_z \leqslant \frac{s_1^{a+\mu-1} - s_1^z}{s_1^{a+\mu-1} - 1}. \tag{8.12}$$

总结以上讨论,我们存如下的

定理. 如果 $\{p_k\}$ 的均值为零，则我们的输光问题的解满足不等式(8.11)；如果 $\{p_k\}$ 的均值不为零，则这个解满足不等式(8.12)。其中 s_1 为(8.6)的不等于 1 的唯一正根，μ 与 ν 分别为使得 $p_k \neq 0$ 的最大与最小的下标。

设 $m = \sum k p_k$ 为一次试验中的期望赢利（或步长的期望值）. 由(8.6)容易看出，当 $m < 0$ 时，$s_1 > 1$；当 $m > 0$ 时，$s_1 < 1$。令 $a \to \infty$，由我们的定理可以推出，在与具有无限赌本的对手进行赌博时，最终输光的概率为 1 的充要条件是 $m \leqslant 0$。

赌博的持续时间可用类似的方法来讨论(参见问题 9)。

9. 问　　题

1. 试就作了如下修改的随机徘徊考虑第 2 节与第 3 节中的输光问题：质点每次向右或向左移动一单位步，或者停留在原来的位置，且对应的概率分别为 α，β，γ $(\alpha + \beta + \gamma = 1)$。(用赌博的话来说，就是赌博可以有平局。)

2. 就原点为弹性壁(如第 1 节所定义)的情况考虑第 2 节与第 3 节中的输光问题. 输光（在原点吸收）概率的差分方程与期望持续时间的差分方程是相同的，但边界条件不同。

3. 设质点每步向右移动两单位或向左移动一单位，其对应的概率分别为 p 与 q $(p + q = 1)$. 如果出发位置为 $z > 0$，试求质点迟早会到达原点的概率。(这是与具有无限赌本的对手进行赌博的输光问题。)

提示：对应于(2.1)的方程有特解 $q_z = 1$ 及两个形如 λ^z 的特解，其中 λ 满足一个二次方程。

4. 在第 8 节中的广义随机徘徊中令[与(8.1)类似] $\rho_z = p_{a-z} + p_{a+1-z} + p_{a+2-z} + \cdots$，并设 $d_{z,n}$ 为赌博恰在第 n 步结束的概率。证明当 $n \geqslant 1$ 时有

$$d_{z,n+1} = \sum_{x=1}^{a-1} d_{x,n}\, p_{x-z},$$

其中 $d_{z,1} = r_z + \rho_z$。因而证明母函数 $d_z(s) = \sum d_{z,n} s^n$ 为线性方程组

$$s^{-1}d_z(s) - \sum_{x=1}^{a-1} d_x(s)\, p_{x-z} = r_z + \rho_z$$

的解．由微分法可以推出，期望持续时间 e_z 为

$$e_z - \sum_{x=1}^{a-1} e_x p_{x-z} = 1$$

的解．

5．在以点 0 与 a 为吸收壁、初始位置为 z 的随机徘徊中，设 $w_{z,n}(x)$ 为质点于第 n 步到达位置 x 的概率．求确定 $w_{z,n}(x)$ 的差分方程与边界条件．

6．（续上）．试给出当点 0 与 a 为反射壁(即 $\delta = 1$ 的弹性壁)时的边界条件．

注意．在以下的问题中，$v_{x,n}$ 是在由原点出发的随机徘徊中，质点于第 n 步到达位置 x 的概率 (6.1)．

7．**映象法**[1]．设 $p = q = 1/2$．考虑在原点具有吸收壁、初始位置为 z 的 $(0, \infty)$ 中的随机徘徊．设 $u_{z,n}(x)$ 为质点于第 n 步到达位置 x 的概率．证明 $u_{z,n}(x) = v_{x-z,n} - v_{x+z,n}$．[提示．证明对应于 (4.1) 的差分方程及适当的边界条件得以满足．]

8．（续上）．如果原点为反射壁，则

$$u_{z,n}(x) = v_{x-z,n} + v_{x+z,n}.$$

9．（续上）．如果随机徘徊限制在 $(0, a)$ 中，且两个壁都是吸收的，则

$$u_{z,n}(x) = \sum_k \{v_{x-z-2ka,n} - v_{x+z-2ka,n}\}, \qquad (9.1)$$

其中的和是对所有的 k（正的或负的）来求的(仅有有限多项异于零)．如果两个壁都是反射的，则用加号代替减号后方程 (9.1) 仍成立．

10．**最大坐标的分布**．在自原点出发的对称自由随机徘徊中，设 $\mathbf{M}_n$ 为质点在时刻 $0, 1, 2, \cdots, n$ 的最大横坐标．利用问题 7 中的公式证明

$$\mathbf{P}\{\mathbf{M}_n = z\} = v_{z,n} + v_{z+1,n}. \qquad (9.2)$$

11．设 $V_x(s) = \sum v_{x,n}s^n$（参看问题 7 前面的注意）．证明当 $x \leqslant 0$ 时 $V_x(s) = V_0(x)\lambda_2^{-x}(s)$，当 $x \geqslant 0$ 时 $V_x(s) = V_0(s)\lambda_1^{-x}(s)$，其中 $\lambda_1(s)$ 与 $\lambda_2(s)$

1) 问题 7—9 是映象法的例子．项 $v_{x-z,n}$ 对应于自由随机徘徊中一个质点，$v_{x+z,n}$ 则对应于一个“象点”．在方程 (9.1) 中有从各种位置出发的象点，它们由在两个边界处的重复反射而得到．在问题 12 与 13 中，我们利用母函数得到关于非对称随机徘徊的一般结果．在差分方程论中，总是把映象法归功于凯尔文．等价的反射原理则通常归功于安德．参见上册第 74 页脚注．

由 (4.8) 定义。此外，$V_0(s) = (1 - 4pqs^2)^{-1/2}$。

注。第 4 节末曾证明，$\lambda_1(s)$ 与 $\lambda_2(s)$ 是初过时间的母函数；以上关系可直接由这个事实推出。

12. 考虑在原点处具有吸收壁、初始位置为 z 的 $(0, \infty)$ 中的随机徘徊，设 $u_{z,n}(x)$ 为质点于第 n 步到达位置 x 的概率，令

$$U_z(s; x) = \sum_{n=0}^{\infty} u_{z,n}(x)s^n. \tag{9.3}$$

利用问题 11 证明 $U_z(s;x) = V_{x-z}(s) - \lambda_2^z(s)\, V_x(s)$。由此推出

$$u_{z,n}(x) = v_{x-z,n} - (q/p)^z \cdot v_{x+z,n}. \tag{9.4}$$

试与问题 7 中的结果相比较，并用组合方法由后者导出 (9.4)。

13. **输光概率的另一公式**。将 (4.11) 展开为几何级数，由此证明

$$u_{z,n} = \sum_{k=0}^{\infty} \left(\frac{p}{q}\right)^{ka} w_{z+2ka,n} - \sum_{k=1}^{\infty} \left(\frac{p}{q}\right)^{ka-z} w_{2ka-z,n},$$

其中 $w_{x,n}$ 由 (5.9) 定义。

14. 如果将第 6 节中的极限过程应用于前一个问题中所给出的 $u_{z,n}$ 的表达式，试证明在长为 Δt 的一个很短的时间间隔内吸收的概率渐近地为[1]

$$\frac{1}{2}\Delta t(\pi D t^3)^{-1/2} e^{-c(ct+z)/D} \sum_{k=-\infty}^{+\infty} (z + 2ka)\, e^{-1/4(z+2ka)^2/Dt}.$$

(提示。应用正态逼近于二项分布。)

15.[2] **输光问题的更新方法**。在具有两个吸收壁的随机徘徊中，令 $u_{z,n}$ 与 $u_{z,n}^*$ 分别为在左壁和右壁处吸收的概率。以适当的解释证明下列两个方程成立：

$$V_{-z}(s) = U_z(s)V_0(s) + U_z^*(s)V_{-a}(s),$$
$$V_{a-z}(s) = U_z(s)V_a(s) + U_z^*(s)V_0(s).$$

通过从这两个方程中解出 $U_z(s)$ 来推导 (4.11)。

16. 设 $u_{z,n}(x)$ 为从 z 出发的质点于第 n 步到达 x 处的概率。利用问题 15 中的记号，证明对于相应的母函数 $U_z(s; x) = \sum u_{z,n}(x)s^n$ 我们有

$$U_z(s; x) = V_{x-z}(s) - U_z(s)\, V_x(s) - U_z^*(s)\, V_{x-a}(s).$$

(不需计算。)

17. (续上)。上个问题中的母函数可用如下方法得到：令

---

1) 这新公式与极限形式 (6.10) 的一致性是西它函数论中的众所周知的事实。

2) 问题 15—17 包含有关于一维随机徘徊的一些主要结果的独立的新推导。

$$U_z(s, x) = V_{x-z}(s) - A\lambda_1^z(s) - B\lambda_2^z(s),$$

选择适当的常数 A, B, 使得当 $z = 0$ 及 $z = a$ 时 $U_z(s; x) = 0$ 的边界条件得以满足. 对于反射壁边界条件为 $U_0(s; x) = U_1(s; x)$ 及 $U_a(s; x) = U_{a-1}(s; x)$.

18.[1] 考虑自原点出发的对称自由随机徘徊. 证明第 r 次返回原点在第 n 步发生的概率等于初过 r 在第 $(n-r)$ 步发生的概率. (提示. 比较两者的母函数.)

19. 用解适当的差分方程的方法证明

$$v_{x,n} = (2\pi)^{-1} 2^n p^{(n+x)/2} q^{(n-x)/2} \int_{-\pi}^{\pi} \cos^n t \cdot \cos tx \cdot dt.$$

进一步推出

$$V_x(s) = (2\pi)^{-1} \left(\frac{p}{q}\right)^{x/2} \int_{-\pi}^{\pi} \frac{\cos tx}{1 - 2(pq)^{1/2} \cdot s \cdot \cos t} dt.$$

20. 在三维对称随机徘徊中质点经过任何特定直线 $x = m$, $y = n$ 无限多次的概率为 1. (提示. 参见问题 1.)

21. 在自原点出发的二维对称随机徘徊中, 质点于第 n 步到达 (x, y) 的概率为

$$(2\pi)^{-2} 2^{-n} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} (\cos\alpha + \cos\beta)^n \cdot \cos x\alpha \cdot \cos y\beta \cdot d\alpha d\beta.$$

证明这个公式并求出三维情况的类似公式. (检验这个表达式满足适当的差分方程.)

22. 在二维对称随机徘徊中, 令 $\mathbf{D}_n^2 = x^2 + y^2$ 为在时刻 n 质点与原点距离的平方. 证明 $\mathbf{E}(\mathbf{D}_n^2) = n$. [提示. 计算 $\mathbf{E}(\mathbf{D}_{n+1}^2 - \mathbf{D}_n)^2$.]

23. 在 d 维随机徘徊中, 质点无限多次返回它先前占据过的位置的概率为 1. [提示: 在每一步移动到新位置的概率最多为 $(2d-1) \div 2d$.]

---

1) 这是第三章第 4 节的定理 3.

第十五章　马尔科夫链

1. 定　　义

直到现在为止，我们主要讨论了独立试验，它可以表述如下．已给一个可能结果 $E_1, E_2, \cdots$，(个数可能有限或无限)的集合，对于每个 E_k，有一个概率 p_k 与之相应；样本序列的概率由乘法性质来确定：$\mathbf{P}\{(E_{j_0}, E_{j_1}, \cdots, E_{j_n})\} = p_{j_0} p_{j_1} \cdots p_{j_n}$．在马尔科夫[1]链理论中，我们考虑一种最简单的推广：允许每次试验的结果依赖于(且仅依赖于)它的直接前一次试验的结果．此时结果 E_k 不再对应于一个固定的概率，但对于每个对 (E_j, E_k) 有一个条件概率 p_{jk} 与之对应；即已知在某次试验中出现 E_j 的条件下，在下次试验中出现 E_k 的概率为 p_{jk}．除 p_{jk} 外，我们还必须给出在初始试验中出现结果 E_k 的概率 a_k．根据 p_{jk} 的意义，对应于二次、三次或四次试验的样本序列的概率必须由下式给定：

$$\mathbf{P}\{(E_j, E_k)\} = a_j p_{jk}, \mathbf{P}\{(E_j, E_k, E_r)\} = a_j p_{jk} p_{kr},$$
$$\mathbf{P}\{(E_j, E_k, E_r, E_s)\} = a_j p_{jk} p_{kr} p_{rs},$$

一般有

$$\mathbf{P}\{(E_{j_0}, E_{j_1}, \cdots, E_{j_n})\} = a_{j_0} p_{j_0 j_1} p_{j_1 j_2} \cdots p_{j_{n-2} j_{n-1}} p_{j_{n-1} j_n}. \tag{1.1}$$

这里初始试验用数字 0 来标号，故标号为 1 的试验是第二次试验．(这个约定是方便的，在上章中就已引进了这个约定而未声明.)

例．(a) 每个马尔科夫链都等价于如下的**罐子模型**．每一个下标用一个罐子来表示，每个罐中各装有带有记号 $E_1, E_2, \cdots$ 的一些球．每个罐的成分是固定的，但各个罐的成分可以不同．在第 j 个罐中抽出带有记号 E_k 的球的概率为 p_{jk}．在开始时(即第

---

1) A. A. Марков (1856—1922).

零次试验)，按概率分布 $\{a_i\}$ 任选一罐，并从该罐中随机地抽出一球，如果该球的记号为 E_j，则下次从第 j 个罐中抽球，如此类推. 显然，按这种手续，序列 $(E_{j_0}, E_{j_1}, \cdots, E_{j_n})$ 的概率由 (1.1) 给出. 我们看出，马尔科夫链的概念并不比罐子模型更一般，但以后将看到，新的记号是更实际和更直观的.

(b) 独立试验当然是以上概型的一个特例，此时对一切 j 有 $p_{jk} = a_k$.

如果 a_k 是在初始(或第零次)试验出现 E_k 的概率，则 $a_k \geqslant 0$ 且 $\sum a_k = 1$. 此外，在 E_j 出现后，下次试验必定出现某个 E_k，故对于一切 j 与 k 必定有

$$p_{j1} + p_{j2} + p_{j3} + \cdots = 1, \quad p_{jk} \geqslant 0. \tag{1.2}$$

我们要证明，对于满足这些条件的任何 a_k 与 p_{jk}，由 (1.1) 来定义对应于 $n+1$ 次试验的样本空间中的概率是合理的. 因为由(1.1)定义的数是非负的，故我们只需证明其和为 1. 首先固定 $j_0, j_1, \cdots, j_{n-1}$ 并将 (1.1) 中的数对所有可能的 j_n 求和. 利用 (1.2)(其中 $j = j_{n-1}$)，我们立即看到，此和为 $a_{j_0}p_{j_0j_1}\cdots p_{j_{n-2}j_{n-1}}$. 因此所有 (1.1) 型数之和不依赖于 n，又因为 $\sum a_{j_0} = 1$，故对于所有的 n，此和等于 1.

定义 (1.1) 形式上依赖于试验次数，但由以上的论证可知，定义 (1.1) 对于所有的 n 是一致的. 例如，为了得到事件"头两次试验的结果为 (E_j, E_k)" 的概率，我们必须固定 $j_0 = j$, $j_1 = k$ 而将 (1.1) 对所有的 $j_2, j_3, \cdots, j_n$ 求和. 我们已经证明此和为 $a_j p_{jk}$，故它与 n 无关. 因此，通常没有必要明确地提到试验的次数；事件 $(E_{j_0}, \cdots, E_{j_r})$ 在所有试验次数大于 r 的样本空间中有相同的概率. 关于独立试验，我们曾多次指出，从数学观点来看，最好是只引入唯一的无限试验序列的样本空间，并把有限次试验的结果看成是一个无限序列的开头部分[1]. 同样的叙述对于马尔科夫链也是正确的. 可惜，无限多次试验的样本空间超出离散概率理论的

---

1) 确切地说，应该看成是开头部分相同的无限序列的一个集合 ——译者注.

范围，而本书第一卷是限于讨论这种理论的.

综上所述，我们的出发点是：

定义. 可能结果为 $E_1, E_2, \cdots$ 的一列试验称为一个马尔科夫链[1]，如果样本序列的概率由 (1.1) 定义，其中 $\{a_k\}$ 为状态 E_k 在时刻 0 的初始概率分布，p_{jk} 为在上一次试验中出现 E_j 的条件下，下一次试验出现 E_k 的条件概率.

现在我们将改变以上的术语，使之与物理应用中的习惯相一致. 我们把"第 n 次试验出现结果 E_k" 说成是在时刻 n 系统处于状态 E_k，并称条件概率 p_{jk} 为 $E_j \to E_k$（由状态 E_j 到状态 E_k）的转移概率.

转移概率可以排成一个转移概率矩阵：

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ p_{31} & p_{32} & p_{33} & \cdots \\ \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdots \end{bmatrix} \tag{1.3}$$

其中第一个足标表示行，第二个足标表示列. 显然 P 是具有非负元素且行和为 1 的方阵. 这种矩阵（有限或无限）称为随机矩阵. 任何随机矩阵都可作为一个转移概率矩阵；连同初始分布，它完全确定一个具有状态 $E_1, E_2, \cdots$ 的马尔科夫链.

在某些特殊情况，将状态从零开始编号比从 1 开始编号更方便一些. 这时要把第 0 行和第 0 列添加到 P 中去.

2. 例　　题

本节中包含很多例子，它们将帮助读者熟悉马尔科夫链的概

1) 这不是标准术语. 这里我们所考虑的仅仅是一类特殊的马尔科夫链，严格地说，在这里及以后各节中，马尔科夫链这一术语应当用"具有不变转移概率"这一短语来加以限制. 实际上，一般类型的马尔科夫链是很少被研究的. 它的定义将在第 10 节中给出，在该节中马尔科夫性将联系一般的随机过程来讨论. 在那里读者还可找到并不构成马尔科夫链的相依试验的例子.

念．为了节省篇幅，以后我们还要参考本节中的某些例子，但并不要求读者将它们背熟．关于古典的洗牌例子见第 9 节．

（a）**假定只有两个状态 $\mathbf{E}_1$ 与 $\mathbf{E}_2$**（也分别称为"成功"与"失败"）．这时矩阵 P 具有如下形式：

$$P=\begin{bmatrix} p & q \\ p' & q' \end{bmatrix}, \quad p+q=p'+q'=1,$$

其中 p, p' 分别为成功接着成功出现和成功接着失败出现的概率．作为一个特殊的例子，设想一个球以速度 ± 1 沿 x 轴运动．在时刻 $1, 2, \cdots$，球以概率 q 改变其运动的方向，而以概率 p 保持其运动方向．如果用 E_1 代表速度 $+1$，E_2 代表 -1，则转移概率矩阵具有以上的形式，其中 $q'=p$，$p'=q$．（这个实验可用大的规则钉板来模拟．）

（b）**具有吸收壁的随机徘徊．** 设可能状态为 $E_0, E_1, \cdots, E_a$，考虑转移概率矩阵

$$P=\begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ q & 0 & p & 0 & \cdots & 0 & 0 & 0 \\ 0 & q & 0 & p & \cdots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdots & q & 0 & p \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix}.$$

从每个"内部"状态 $E_1, \cdots, E_{a-1}$ 只可能转移右边或左边的邻近状态（相应的概率为 $p_{i,i+1}=p$，$p_{i,i-1}=q$）．然而，从 E_0 或 E_a 中任一个不可能转移到任何其它状态；系统可以从一个状态转移到另一个状态，但一旦到达 E_0 或 E_1，则系统就永远停留在那里．显然，这个马尔科夫链仅在术语上与上章所讨论的在 0 与 a 处具有吸收壁的随机徘徊不同．在那里，随机徘徊从区间中的一个固定点 z 出发．用马尔科夫链的术语来说，这相当于选择初始分布，使得 $a_z=1$（因而当 $x\neq z$ 时 $a_x=0$）．如果我们随机地选取初始状态，则初始分布为

$$a_k = (a+1)^{-1},\ k = 0, 1, \cdots, a.$$

（c）**弹性壁**. 下面我们考虑一个矩阵，它与前面的矩阵仅在第 1 行[1]和第 $a-1$ 行不同. 选取 $0 \leqslant \delta_0 \leqslant 1$ 和 $0 \leqslant \delta_a \leqslant 1$，并令

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ (1-\delta_0)q & \delta_0 q & p & 0 & \cdots & 0 & 0 & 0 \\ 0 & q & 0 & p & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & p & 0 \\ 0 & 0 & 0 & 0 & \cdots & q & \delta_a p & (1-\delta_a)p \end{bmatrix}.$$

除了由 E_1 到 E_0 的概率为 $(1-\delta_0)q$，停留在 E_1 的概率为 $\delta_0 q$ 以及关于 E_{a-1} 的类似性质外，这里的转移概率是和上例相同的. 当 $\delta_0 = \delta_a = 0$ 时，我们的矩阵和上例中的矩阵完全相同. 如果 $\delta_0 = \delta_a = 1$，则 E_0 与 E_a 是不能到达的；系统从一个内部状态出发，将从一个状态到达另一个状态，但永远不能到达 E_0 或 E_a. 用随机徘徊的术语来说，后面的情形对应于反射壁（参见第十四章）. 用赌博的语言来说，系统的状态表示一个赌徒的赌本，在这种赌博中，两赌徒的赌本之和为 a. 当第一个赌徒输掉最后一元时，他的对手将这一元退还给他的概率为 δ_0，而赌博结束的概率则为 $1-\delta_0$. 具有两个反射壁的赌博永远不会结束.

（d）**循环随机徘徊**. 仍设可能状态为 $E_1, E_2, \cdots, E_a$，但将它们按循环次序排列，于是 E_a 具有邻近状态 E_{a-1} 与 E_1. 和以前一样，如果系统只能到达右边或左边的邻近状态，则矩阵 P 的第一行为 $(0, p, 0, 0, \cdots, 0, q)$，最后一行为 $(p, 0, 0, 0, \cdots, 0, q, 0)$，其余各行则和例（b）中的矩阵一样.

更一般地，我们可以允许在任何两个状态之间发生转移. 设

1）根据上节末的约定，矩阵在通常意义下的第一行现在称为第 0 行，故此处所说的第一行就是通常意义下的第二行. ——译者注.

$q_0, q_1, \cdots, q_{a-1}$ 分别表示固定不动或向右移动 $1, 2, \cdots, a-1$ 个单位的概率(这里,向右移动 k 单位与向左移动 $a-k$ 单位是一样的). 于是 P 是一个循环矩阵

$$P = \begin{bmatrix} q_0 & q_1 & q_2 & \cdots & q_{a-2} & q_{a-1} \\ q_{a-1} & q_0 & q_1 & \cdots & q_{a-3} & q_{a-2} \\ q_{a-2} & q_{a-1} & q_0 & \cdots & q_{a-4} & q_{a-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ q_1 & q_2 & q_3 & \cdots & q_{a-1} & q_0 \end{bmatrix}.$$

如果 $q = p$, $q_{a-1} = q$ 且当 $1 < k < a-1$ 时, $q_k = 0$, 则这个随机徘徊就变成本例开头所讨论的简单情况. [在第十六章例 (2.d) 中将续继讨论这个问题.]

(e) 自由随机徘徊. 一个自由的一维随机徘徊是一个马尔科夫链,但最自然的是将状态排成一个双重无限的序列($\cdots E_{-2}, E_{-1}, E_0, E_1, E_2, \cdots$). 为了将转移概率写成习惯的形式,我们必须将状态重新加以排列. 例如,对于次序($E_0, E_1, E_{-1}, E_2, E_{-2}, \cdots$), P 的第一行变成 $(0, p, q, 0, 0, \cdots)$, 第二行变成 $(q, 0, 0, p, 0, 0, \cdots)$, 等等. 遗憾的是,这时不再具有自然的对称性,且公式变得不能令人满意. 在二维的场合,情况则更坏,在这种场合,本章的方法不便于导出明显的公式,但一般定理是适用的.

(f) 扩散的爱伦弗斯特 (Ehrenfest) 模型. 我们再次考虑具有 $a+1$ 个状态 $E_0, E_1, \cdots, E_a$, 且只可能向右边或左边的邻近状态转移的链;但此时我们令 $p_{j,j+1} = 1 - j/a$, $p_{j,j-1} = j/a$, 故

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ a^{-1} & 0 & 1-a^{-1} & 0 & \cdots & 0 & 0 \\ 0 & 2a^{-1} & 0 & 1-2a^{-1} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & a^{-1} \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$$

这个链有两个有趣的物理解释．为了讨论统计力学中的各
循环问题，爱伦弗斯特[1]考虑了如下的假想实验：设有 a 个分子
布在两个容器 A 与 B 中．在时刻 n 随机地选取一个分子并把它
原来所在的容器移到另一个容器中去．系统的状态由 A 中分子
数目来确定．假定在某一时刻，容器 A 中恰有 k 个分子．在下
试验中，系统变成 E_{k-1} 或 E_k，这要看分子是自 A 中选出还是自
中选出而定；相应的概率为 k/a 与 $(a-k)/a$，故我们的链描述
爱伦弗斯特的实验．然而，这个链也可解释为在中心力作用下
扩散，即一个向右移动的概率随着质点的位置而改变的随机徘徊
如果 $j < a/2$，则 j 向右移动比向左移动的可能性更大；如果 j
$a/2$，则由 j 向左移动比向右移动的可能性更大．这意味着，质
有朝向 $x = a/2$ 移动的趋势，这对应于一个与距离成正比增加
吸引弹性力．（爱伦弗斯特模型曾在第五章例 (2.c) 中讨论过；
见本章例 (6.a) 及问题 12．）

（**g**）**占位问题**．在第一章中我们考虑过将球随机地放入 a
盒中的问题．设系统的状态由非空盒的数目来确定．如果有 j
盒是非空的，则下次试验中球被放入一个空盒的概率为 $(a-j)/$
故这个实验可用一个马尔科夫链来描述，其转移概率为 $p_{jj} = j/$
$p_{j,j+1} = (a-j)/a$，对于 j, k 的所有其它组合 $p_{j,k} = 0$．初始
布（所有盒都是空的）为 $p_0 = 1$，当 $1 \leqslant k \leqslant a$ 时 $p_k = 0$．［参
第十六章例 (2.e)．］

（**h**）**成功连贯**．在伯努利试验序列中，如果第 n 次试验的
果为失败，则我们说在时刻 n 我们观察到状态 E_0；如果最后一
失败出现在第 $n-k$ 次试验（第 0 次试验按失败计），则说在第 n
试验我们观察到状态 E_k．换句话说，E_k 的下标 k 表示截至第
次试验为止的不间断的成功序列的长度．显然，以上所讨论的
一个马尔科夫链，其中仅仅转移 $E_k \to E_0$ 与 $E_k \to E_{k+1}$ 是可能的

1) 见 [29] 及 [30]．更复杂的讨论（所用的方法本质上与本书第十六章中的方
等价）见 [31]．又见 [32]．

于是转移矩阵具有如下形式:

$$P=\begin{bmatrix} q & p & 0 & 0 & 0 & \cdots \\ q & 0 & p & 0 & 0 & \cdots \\ q & 0 & 0 & p & 0 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \end{bmatrix}.$$

(i) **循环事件.** 上面的例子是一个更有趣的马尔科夫链的特殊情况. 设 $\mathscr{E}$ 为任一循环事件, 其循环分布为 $\{f_n\}$. 习惯上, 我们总是说第 0 次试验 $\mathscr{E}$ 是出现的. 如果 $\mathscr{E}$ 在第 n 次试验出现, 则我们说在时刻 n 系统处于状态 E_0; 如果 $\mathscr{E}$ 最后一次出现是在第 $n-k$ 次试验, 则我们说系统处于状态 E_k. (也可以说, 我们所论及的是负方向的等待时间.) 和上个例子中一样, 如果在第 n 次试验系统处于状态 E_k, 则显然在下次试验中系统只可能转移到状态 E_0 (如果 $\mathscr{E}$ 出现) 或 E_{k+1}. 令

$$s_k=f_1+\cdots+f_k,\quad q_k=\frac{f_{k+1}}{1-s_k},$$

$$p_k=1-q_k=\frac{1-s_{k+1}}{1-s_k}. \tag{2.1}$$

注意, E_k 意味着 $\mathscr{E}$ 的等待时间超过 k, 且在这个假设下, 下次试验出现 $\mathscr{E}$ 的概率等于 q_k. 因此转移 $E_k\to E_{k+1}$ 与 $E_k\to E_0$ 的概率分别为 p_k 与 q_k. 一个典型的样本序列具有形状 $E_0E_0E_1E_2E_3E_0E_1E_0E_1E_2E_0E_0$ (第一个 E_0 表示第 0 次试验). 这里等待时间依次为 1, 4, 2, 3, 1, 而这个样本序列的概率则为 $f_1f_4f_2f_3f_1$. 现在

$$f_1f_4f_2f_3f_1=q_0p_0p_1p_2q_3p_0q_1p_0p_1q_2q_0, \tag{2.2}$$

这与马尔科夫链中的概率公式 (1.1) 是相符的. 以上的推理适用于所有序列, 于是我们得到一个马尔科夫链, 其转移概率矩阵为

$$P = \begin{bmatrix} q_0 & p_0 & 0 & 0 & 0 & \cdots \\ q_1 & 0 & p_1 & 0 & 0 & \cdots \\ q_2 & 0 & 0 & p_2 & 0 & \cdots \\ q_3 & 0 & 0 & 0 & p_3 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \end{bmatrix}.$$

[在例(6.c)中将继续讨论这个问题.]

(j) **序贯抽样.** 如我们已在第十四章第8节中所看到的，在序贯抽样中遇到如下问题. 设 $\mathbf{S}_n = \mathbf{X}_1 + \cdots + \mathbf{X}_n$，此处 $\mathbf{X}_\nu$ 是仅取整数值的独立同分布的随机变量，其共同分布为 $\{p_k\}$，$k=0$，±1，±2，$\cdots$. 对于预先指定的 $z>0$，$b>0$，存在一个最小的 n，使得 $\mathbf{S}_n \geqslant b$ 或 $\mathbf{S}_n < -z$. 这个 n 当然是一个随机变量，我们所关心的是它的分布及随机事件 $\mathbf{S}_n \leqslant -z$ 与 $\mathbf{S}_n \geqslant b$ 的概率.

这个问题可用具有状态 $0, 1, 2, \cdots, b+z$ 的马尔科夫链来描述. 设 $a = b+z-1$，并选 z 作为初始状态. 如果 $z+\mathbf{S}_n = x$ 并且和 $z+\mathbf{S}_1, \cdots, z+\mathbf{S}_{n-1}$ 中没有一个是 $\leqslant 0$ 或 $> a$，则我们说在时刻 n，系统处于状态 $x(x=1, 2, \cdots, a)$；在同样的附加条件下，如果 $z+\mathbf{S}_n \leqslant 0$，则我们说系统处于状态 0，如果 $z+\mathbf{S}_n \geqslant a+1$，则说系统处于状态 $a+1$. 一旦系统变成两个极限状态 0 与 $a+1$ 中的一个，则它将永远停留在那里(即，我们令 $p_{0,0} = p_{a+1,a+1} = 1$). 转移概率矩阵为

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ r_1 & p_0 & p_1 & p_2 & \cdots & p_{a-1} & \rho_1 \\ r_2 & p_{-1} & p_0 & p_1 & \cdots & p_{a-2} & \rho_2 \\ r_3 & p_{-2} & p_{-1} & p_0 & \cdots & p_{a-3} & \rho_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ r_a & p_{-a+1} & p_{-a+2} & p_{-a+3} & \cdots & p_0 & \rho_a \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix},$$

其中

$$r_k = p_{-k} + p_{-k-1} + p_{-k-2} + p_{-k-3} + \cdots$$

及

$$\rho_k = p_{a-k+1} + p_{a-k+2} + \cdots.$$

作为一个例子，我们考虑巴特基的二重抽样检查方案．为了检查一批货物，抽出大小为N的样本，并进行全面检查．假定样本是统计独立的，且每个样本中次品的个数具有相同的二项分布．考虑到每个样本中有一个次品的情况，故我们令$\mathbf{X}_k + 1$等于第k个样本中次品的个数．于是当$k \geqslant 0$时有

$$p_k = \binom{N}{k+1} p^{k+1} q^{N-k-1} \tag{2.3}$$

及 $p_{-1} = q^N$，$p_x = 0 (x < -1)$．检验的程序规则如下：抽出一个初始样本，如果其中无次品，则整批货物就被接受；如果废品个数超过 a，则这批货物就被拒绝．当这两种情况中的任何一种发生时，检查就不再进行，因此也就没有马尔科夫链．然而，如果次品数z在范围$1 \leqslant z \leqslant a$内，则只要链的状态还在1与$a$之间，抽样就按上述方式一直进行下去．链的状态迟早会变成0（在这种情况，货物被接受），或变成$a+1$（在这种情况，货物被拒绝．）

(k) 遗传学中的一个例子[1]．在每一代中选择N个个体而构成总体（于是总体的大小保持不变）．在每一代中，一个取A与a两种形式的特殊基因有$2N$个代表．如果在第n代中A出现j次，则a出现$2N-j$次．在这种情况，我们说在时刻n总体处于状态$j(0 \leqslant j \leqslant 2N)$．假定交配是随机的，则下一代的成分由$2N$次伯努利试验所确定，其中$A$型基因具有概率$j/2N$．因此我们得到一个马尔科夫链，其转移概率为

$$p_{jk} = \binom{2N}{k} \left(\frac{j}{2N}\right)^k \left(1 - \frac{j}{2N}\right)^{2N-k}.$$

1) 费歇与赖特 (S. Wright) 详细地讨论过这个问题．用马尔科夫链来阐述是马兰考脱 (G. Malécot) 在 [33] 中给出的．

[参见例(8.c).]

(l) **一个繁殖问题.** 在所谓兄-妹交配中,使两个个体进行交配,在它们的直接后代中随机选取两个异性个体,并使之交配. 这个过程无限地继续下去. 当每个亲本具有三种遗传型 AA, Aa, aa 时,我们必须区别亲本的如下六种组合: $E_1 = AA \times AA$, $E_2 = AA \times Aa$, $E_3 = Aa \times Aa$, $E_4 = Aa \times aa$, $E_5 = aa \times aa$, $E_6 = AA \times aa$. 利用第五章中的法则,容易看到,在这种情况转移概率矩阵为

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 \\ \frac{1}{16} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{16} & \frac{1}{16} \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

[在习题中将继续讨论这个问题;在第十六章例(4.b)中给出了完备的论述.]

3. 高阶转移概率

由 E_j 经恰好 n 步而转移到 E_k 能有各种途径: $E_j \to E_{j_1} \to E_{j_2} \to \cdots \to E_{j_{n-1}} \to E_k$. 在已知系统处于状态 E_j 的情况下系统经过这个特殊途径的条件概率为 $p_{jj_1} p_{j_1j_2} \cdots p_{j_{n-1}k}$. 所有可能途径的这种对应表达式之和就是已知系统在时刻 r 处于状态 E_j 的情况下,在时刻 $r+n$ 系统处于状态 E_k 的概率. 我们用 $p_{jk}^{(n)}$ 来表示这个概率.

我们特别有 $p_{jk}^{(1)} = p_{jk}$ 及

$$p_{jk}^{(2)} = \sum_{\nu} p_{j\nu} p_{\nu k}. \tag{3.1}$$

应用归纳法容易求得递推公式

$$p_{jk}^{(n+1)} = \sum_{\nu} p_{j\nu} p_{\nu k}^{(n)}; \tag{3.2}$$

进一步对 m 进行归纳,可以得到如下更一般的公式

$$p_{jk}^{(m+n)} = \sum_{\nu} p_{j\nu}^{(m)} p_{\nu k}^{(n)}. \tag{3.3}$$

这个方程反映了如下的一个简单事实：前 m 步系统由 E_j 到达某个中间状态 E_ν，后 n 步则从 E_ν 到达 E_k. 恒等式(3.3)是马尔科夫链的特征. 对于更一般的过程(参见第10节),一个类似的方程成立,但最后的因子不仅依赖于 ν 与 k 而且也依赖于 j.

用由 p_{jk} 构成矩阵 P 的同样方法，我们将 $P_{jk}^{(n)}$ 排成一个矩阵，并用 P^n 来表示. 方程(3.2)表明,为求 P^{n+1} 的元素 $p_{jk}^{(n+1)}$，我们只要把 P 的第 j 行的元素乘以 P^n 的第 k 行的对应元素,并把所有这些乘积加起来. 这种运算称为矩阵 P 与 P^n 的行对列乘法,在符号上用等式 $P^{n+1} = PP^n$ 来表示. 因此我们称 P^n 为 P 的 n 次幂;等式(3.3)则表示结合律 $P^{m+n} = P^m P^n$.

为了使等式(3.3)对一切 $n \geqslant 0$ 成立,我们定义 $p_{jk}^{(0)}$ 如下:

$p_{jj}^{(0)} = 1$; 当 $j \neq k$ 时 $P_{jk}^{(0)} = 0$.

例. (a) 在独立试验这种最简单的情况，P 的所有各行都是一样的,故不用计算就可看出,对一切 n 有 $P^n = P$.

(b) 在成功连贯[例(2.h)]中，n 步转移概率可以直接求出. 例如,经过三步后，系统由 E_k 仅能到达 E_{k+3}，E_0，E_1，E_2，对应的概率显然为 p^3，q，qp，qp^2. 于是

$$P^2 = \begin{bmatrix} q & qp & p^2 & 0 & 0 & \cdots \\ q & qp & 0 & p^2 & 0 & \cdots \\ q & qp & 0 & 0 & p^2 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \end{bmatrix},$$

$$P^3 = \begin{bmatrix} q & qp & qp^2 & p^3 & 0 & 0 & \cdots \\ q & qp & qp^2 & 0 & p^3 & 0 & \cdots \\ q & qp & qp^2 & 0 & 0 & p^3 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \end{bmatrix}.$$

在这种情况，P^n 显然收敛于第 k 列元素等于 qp^k 的矩阵.

绝对概率. 再次令 a_j 表示系统在时刻 0 处于状态 E_j 的概率. 于是在时刻 n 系统处于状态 E_k 的(无条件)概率为

$$a_k^{(n)} = \sum_j a_j p_{jk}^{(n)}. \tag{3.4}$$

通常，我们令过程从一个固定状态 E_i 开始，即我们取 $a_i = 1$. 在这种情况，我们有 $a_k^{(n)} = p_{ik}^{(n)}$.

直观上我们感到，初始状态对时刻 n 的概率分布的影响应当随着 n 的增大而逐渐减弱，因此当 n 很大时，分布 (3.4) 应当几乎和初始分布 $\{a_j\}$ 无关. 如果 $P_{jk}^{(n)}$ 收敛于一个不依赖于 j 的极限，即 P^n 收敛于一个各行相同的矩阵(如上例中那样)，则上述情况发生，我们将看到，通常是如此的，但我们必须再一次考虑由于周期性而引起的麻烦的例外情况.

4. 闭包和闭集

如果存在某个 $n \geq 0$，使得 $p_{jk}^{(n)} > 0$(即，由 E_j 到达 E_k 的概率为正，其中包括 $E_j = E_k$ 的情形)，则我们称由 E_j 能到达 E_k. 例如，在自由随机徘徊中，由每一状态都能到达任何其它状态，但由吸收壁不能到达任何其它状态.

定义. 设 C 是状态的一个集合. 如果从 C 内的任何状态 E_j 不能到达 C 外的任何状态，则称 C 是闭的. 包含 C 的最小闭集称为 C 的闭包.

如果单个状态 E_k 构成一个闭集，则称 E_k 为吸收的.

一个马尔科夫链称为是不可约的，如果除了所有状态的集合外，不存在任何其它闭集.

显然 C 为闭集的充要条件是，当 j 在 C 中而 k 在 C 外时恒有 $p_{jk}=0$；因为在这种情况下，由(3.2)可知，对于一切 n 有 $P_{jk}^{(n)}=0$. 于是我们有如下显然的

定理. 在 P^n 中如果将对应于闭集 C 外状态的所有行和列都删去，则剩下一个随机矩阵，其中基本关系(3.2)与(3.3)仍成立.

这意味着，我们有一个定义在 C 上的马尔科夫链，且这个子链可以不涉及所有其它状态而独立地予以研究.

当且仅当 $p_{kk}=1$ 时，状态 E_k 是吸收的，在这种情况，上述定理中的矩阵缩成一个单一的元素. 状态 E_j 的闭包是能由它到达的所有元素(包括 E_j 在内)的集合. 这个附注可以改述为如下有用的

准则. 一个链为不可约的充要条件是由每个状态能到达任何其它状态.

例. 为了求出所有闭集，只要知道 p_{jk} 中哪些为0哪些为正就够了. 因此，我们用 * 表示正元素，并考虑一个典型矩阵，如

$$P=\begin{bmatrix} 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & * \\ 0 & * & * & 0 & * & 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & 0 \\ * & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 & * & * & 0 & 0 \\ 0 & 0 & * & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & * \end{bmatrix}.$$

在第五行中只在第五列处出现一个 *，所以 $p_{55}=1$，故 E_5 是吸收的. 第三行和第八行各仅有一个正元素，故显然 E_3 和 E_8 构成一个闭集. 从 E_1 可能到达 E_4 与 E_9，而从那里仅能到达 E_1, E_4, E_9. 故这三个状态 E_1, E_4, E_9 构成另一闭集.

很清楚，P 的复杂性主要是由于记号不方便而造成的．现在让我们将状态重新编号如下：

$$E'_1 = E_5;\ E'_2 = E_3;\ E'_3 = E_8;$$
$$E'_4 = E_1;\ E'_5 = E_9;\ E'_6 = E_4;$$
$$E'_7 = E_2;\ E'_8 = E_7;\ E'_9 = E_6.$$

将矩阵 P 的元素按同样的方式重新加以排列，则 P 变成如下形式：

$$P' = \begin{bmatrix} * & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & * & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * & * & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * & * & 0 & 0 & 0 \\ 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & 0 \\ * & * & 0 & 0 & * & 0 & * & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & * & 0 & 0 \end{bmatrix}.$$

在这种形式下，(E'_1)，(E'_2, E'_3) 与 (E'_4, E'_5, E'_6) 为闭集是很明显的．从 E'_7 可以到达这三个闭集中的任一个，故 E'_7 的闭包是由状态 $E'_1, E'_2, E'_3, E'_4, E'_5, E'_6, E'_7$ 所构成的集合．由 E'_8 可以到达 E'_7 与 E'_9，因而可以到达每个状态[1]；故 E'_8 的闭包由所有九个状态所组成．由 E'_9 能到达 E'_7（因而能到达 E'_7 的闭包中的各个状态），但不能到达 E'_8，故 E'_9 的闭包是除 E'_8 外的所有状态的集合[2]．

依次删去 (E'_1)，(E'_2, E'_3) 与 (E'_4, E'_5, E'_6) 这三个闭集之外的所有行与列，我们得到三个随机子矩阵

$$[\,*\,] \quad \begin{bmatrix} 0 & * \\ * & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & * & * \\ 0 & * & * \\ * & 0 & 0 \end{bmatrix}, \qquad (4.1)$$

且 P' 不包含与这三个矩阵不相重叠[3]的其它随机子矩阵．

1）原书为闭集，不贴切——译者注．

2）这句话是译者改写的 ——译者注．

3）原书无“不相重叠”的限制，这是不对的；因为闭集不止三个．——译者注

建议读者求出第 2 节各例的矩阵中的闭集和吸收状态.

5. 状态的分类

考虑一个任意的固定状态 E_j，并假定在初始时刻系统处于状态 E_j. 每当系统经过 E_j 时，过程重新和以前一样从头开始. 所以返回 E_j 显然是一个如第十三章所定义的循环事件. 如果系统从另一状态 E_i 开始，则经过 E_j 变成一个如第十三章第 5 节所定义的延迟了的循环事件. 因此，马尔科夫链只不过是循环事件的一个特殊情况；新的特征仅仅在于我们现在是同时考虑很多循环事件.

每个状态的特征可由其循环时间的分布 $\{f_j^{(n)}\}$ 来刻划. 这里 $f_j^{(n)}$ 是在时刻 n 第一次返回 E_j 的概率. $f_j^{(n)}$ 可由 $p_{jj}^{(n)}$ 用如下显然的递推关系来计算[1]：

$$\begin{gathered} f_j^{(1)} = p_{jj}, \quad f_j^{(2)} = p_{jj}^{(2)} - f_j^{(1)} p_{jj}, \quad \cdots, \\ f_j^{(n)} = p_{jj}^{(n)} - f_j^{(1)} p_{jj}^{(n-1)} - f_j^{(2)} p_{jj}^{(n-2)} - \cdots - f_j^{(n-1)} p_{jj}, \end{gathered} \tag{5.1}$$

这些公式当然仅仅是第十三章关于循环事件的基本关系 (3.1) 的一个特殊情形. 和

$$f_j = \sum_{n=1}^{\infty} f_j^{(n)} \tag{5.2}$$

为从 E_j 出发系统迟早要返回 E_j 的概率. 如果 $f_j = 1$，则称状态 E_j 是**常返的**；在这种情况，平均循环时间为

$$\mu_j = \sum_{n=1}^{\infty} n f_j^{(n)}. \tag{5.3}$$

如果 $\mu_j = 0$，则我们称 E_j 为**消极常返**的，或称它为**零状态**.

如果系统从 E_i 出发，则直到第一次经过 E_j 的等待时间具有分布 $f_{ij}^{(n)}$，此处

1）这些关系式表明，在时刻 n 第一次回到 E_j 的概率等于在时刻 n 返回的概率减去在某个时刻 $\nu = 1, 2, \cdots, n-1$ 第一次返回且接着又在时刻 n 返回的概率。用第十三章 (3.1) 中的记号来表示时，我们有 $p_{jj}^{(n)} = u_n$，$f_j^{(n)} = f_n$.

$$f_{ij}^{(1)} = p_{ij}, \quad f_{ij}^{(n)} = p_{ij}^{(n)} - \sum_{\nu=1}^{n-1} f_{ij}^{(n-\nu)} p_{jj}^{(\nu)}. \tag{5.4}$$

这个方程并不是马尔科夫链所特有的，而是对任意的延迟了的循环事件都成立. 显然，如果 E_j 不能由 E_i 到达，则对于一切 n 有 $f_{ij}^{(n)} = 0$. 一般地，

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)} \tag{5.5}$$

是由 E_i 出发，系统迟早要到达 E_j 的概率.

现在我们可将第十三章第 3 节与第 5 节中所证明的事实综述如下：

(1) 如果 $f_j < 1$，则状态 E_j 是非常返的. 对此，充要条件为 $\sum_{n=1}^{\infty} p_{jj}^{(n)} < \infty$. 在这种情况，对于每个 i，$\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$ 显然成立[1].

(2) 如果 $f_j = 1$ 而平均循环时间 $\mu_j = \infty$，则 E_j 是常返零状态. 对此，充要条件是 $\sum_{n=1}^{\infty} p_{jj}^{(n)} = \infty$ 而 $p_{jj}^{(n)} \to 0$. 在这种情况[1]，

$$\text{当 } n \to \infty \text{ 时}, \; p_{ij}^{(n)} \to 0 \tag{5.6}$$

对每个 i 成立.

(3) 如果当 n 不能被正整数 $t > 2$ 除尽时，恒有 $p_{jj}^{(n)} = 0$，且 t 是具有这种性质的数中的最大者(即，除非在第 t, $2t$, $3t$, $\cdots$ 等步上，系统不可能返回 E_j)，则 E_j 具有周期 t.

(4) 如果 E_j 是常返和非周期的，则[1]

$$\text{当 } n \to \infty \text{ 时}, \; p_{ij}^{(n)} \to \mu_j^{-1} f_{ij}, \tag{5.7}$$

特别，

$$\text{当 } n \to \infty \text{ 时}, \; p_{jj}^{(n)} \to \mu_j^{-1}. \tag{5.8}$$

(如果 E_j 是零状态，则令 $\mu_j^{-1} = 0$.)

(5) 如果 E_j 是常返的且有周期 t，则 (5.8) 用下式来代替：

$$\text{当 } n \to \infty \text{ 时}, \; p_{jj}^{(nt)} \to t\mu_j^{-1}. \tag{5.9}$$

---

1) 这可直接由 (5.4) 得到，但它实际上是第十三章第 5 节中的定理的特殊情况.

非周期的常返非零状态称为遍历的[1].

例. (a) 考虑第 4 节的例子中的矩阵 P' (略去了记号“$'$”). 状态 E_1 是常返的(因为它是吸收的). 从 E_2, 系统必定变成 E_3, 而且又从那里回到 E_2. 所以 E_2 和 E_3 是常返状态, 其周期为 2, 平均循环时间亦为 2. 状态 E_4, E_5, E_6 构成一个闭子集, 且它们之间的转移规律由 (4.1) 中最后一个矩阵来表示,由此易知, 这些状态都是常返和非周期的. (稍后我们就将看到,在有限马尔科夫链中不可能有常返零状态.)

从 E_7 到达这些闭集之一是可能的,到达后系统就永远停留在其中;故 E_7 是非常返的. 由 E_9 系统能到达 E_7, 而且由 E_7 不可能回到 E_9; 故 E_9 也是非常返的. 最后,由 E_8 出发,系统迟早将到达 E_7 或 E_9, 而由 E_7 或 E_9 不可能返回 E_8, 故 E_8 是非常返的.

(b) 在自由随机徘徊中[例 (2.e)], 如果 $p=q$, 则所有状态都是常返的;如果 $p \neq q$, 则所有状态都是非常返的[见例 (8.d)].

要判定一个状态是否为常返状态通常是不容易的, 且 $\sum p_{jj}^{(n)}$ 敛散性的判别准则通常也难以应用. 一个较好的准则将在下节的定理中给出.

设 E_j 是一个固定的常返状态, E_k 是能由它到达的某个状态. 此外,设N 为由 E_j 到 E_k 的最短的可能途径的长度,并令 $p_{jk}^{(N)}=\alpha>0$. 由 E_k 到 E_j 的概率必定为正,因为不然的话,系统不能回到 E_j 的概率至少为 α, 于是 $f_j \leqslant 1-\alpha<1$, 而这与 E_j 为常返的假设矛盾. 故存在一个指标 M, 使得 $p_{kj}^{(M)}=\beta>0$. 现在, 对于任意的 n 我们显然有

$$p_{jj}^{(n+N+M)} \geqslant p_{jk}^{(N)} p_{kk}^{(n)} p_{kj}^{(M)}=\alpha\beta \cdot p_{kk}^{(n)} \tag{5.10}$$

及

$$p_{kk}^{(n+N+M)} \geqslant p_{kj}^{(M)} p_{jj}^{(n)} p_{jk}^{(N)}=\alpha\beta \cdot p_{jj}^{(n)}. \tag{5.11}$$

---

1) 遗憾的是,没有被普遍接受的术语. 在第一版中, 常返状态被称为循环状态,但这个名称易与循环事件相混淆. 柯尔莫果洛夫称非常返状态为非本质的, 但新的研究表明, 理论和实践的主要兴趣都集中于非常返状态. 遍历这一术语(它与“常返、非零、非周期”同义)是相当普遍地被接受的, 但“正”状态是现存的其它术语之一,而且有时“遍历”被看成是与常返同义的术语.

由以上二式可知，序列 $p_{jj}^{(n)}$ 与 $p_{kk}^{(n)}$ 具有相同的渐近状态，由此我们能导出重要的结论．首先，我们已假定 E_j 是常返的，故级数 $\sum p_{jj}^{(n)}$ 发散．由 (5.11) 可知，$\sum p_{kk}^{(n)}$ 也发散，故 E_k 必为常返状态．如果 $p_{jj}^{(n)} \to 0$，则也有 $p_{kk}^{(n)} \to 0$，反之亦然．最后，设 E_j 有周期 t．因为经 $N+M$ 步回到 E_j 是可能的，故 $N+M$ 必为 t 的倍数．于是由 (5.10) 与 (5.11) 可知，E_j 与 E_k 必有相同的周期．

因此，我们看到，从常返状态仅能到达常返状态，且它们都具有相同的类型：或者它们都是零状态，或者都是遍历的，或者都是具有相同周期的非零状态．一个常返状态的闭包 C 是一个不可约集，它的子矩阵定义一个以 C 为状态集的马尔科夫链，它可以不依赖于其余状态而被加以研究．于是我们已证明了如下重要的

定理．在不可约马尔科夫链中，所有的状态属于相同的类：或者它们都是非常返的，或者都是常返零状态，或者都是常返非零状态．在每种情况，它们有相同的周期．而且每个状态能到达每个其它状态．

在每个链中，常返状态能唯一地分解为若干个闭集 C_1，C_2，$\cdots$，使得从一个给定闭集 C_ν 中的任一状态能够到达该集中的任一状态而不能到达其它状态．所有属于同一闭集 C_ν 的状态必定是同类的．

除闭集 C_ν 外，链中通常还包含非常返状态，从这些状态能到达闭集 C_ν 中的状态(但反之不成立)．

这个定理有如下有趣的

推论．在有限马尔科夫链中不存在零状态，也不可能所有状态都是非常返状态．

证．考虑不可约链就够了．如果所有状态都非常返状态或都是零状态，则对于每一固定的对 j, k，当 $n \to \infty$ 时我们有 $p_{jk}^{(n)} \to 0$．P^n 的每一行趋向于零，而行和等于 1．显然在有限多项的情况这是不可能的，于是我们得出结论：在有限不可约链中既不存在非常返状态也不存在零状态．

由以上的讨论可知，把状态适当地重新编号后(如我们在第 4

节的例子中所做的那样),对应于链(为叙述方便起见,设这个链有两个闭集 C_1 与 C_2 及一些非常返状态)的矩阵 P 能够写成如下的分块矩阵的形式:

$$P=\begin{bmatrix} P_1 & 0 & 0 \\ 0 & P_2 & 0 \\ A & B & C \end{bmatrix}, \tag{5.12}$$

其中 P_1 与 P_2 是两个闭集中的转移概率矩阵. 于是矩阵 P^n 与 P 有相同的形状, 只是要把 P_1, P_2, C 换成 P_1^n, P_2^n, C^n (而 A 与 B 则要换成更复杂的矩阵, 这将在第 8 节中来研究). 注意, P_1, P_2 与 C 是方阵,而 A 与 B 如同在例题中一样,可以是长方阵.

6. 不可约链的遍历性

在本节中,我们仅讨论非周期链;如同在一般的循环事件中一样,对于周期链只要作一些简单的修改,但要确切地陈述则相当麻烦.

定义. 概率分布 $\{v_k\}$ 称为平稳的,如果

$$v_j=\sum_i v_i p_{ij}. \tag{6.1}$$

如果初始分布 $\{a_k\}$ 碰巧是平稳的,则绝对概率 $\{a_k^{(n)}\}$ 不依赖于时刻 n, 即 $a_k^{(n)}=a_k$. 如果我们设想有大量过程同时进行,则平稳性的物理意义就变得很明显. 假如,设有 N 个质点独立地作同一类型的随机徘徊. 在时刻 n, 处于状态 E_k 的质点的期望数为 $Na_k^{(n)}$. 如果初始分布是平稳的,则这些期望数保持不变,于是我们观察到(如果 N 很大因而可以应用大数定律的话)一个宏观的平衡状态, 它是由于大量相反方向的转移造成的. 物理学中大多数统计平衡属于这种类型;即,它们完全是由于对很多独立质点的同时观察的结果. 典型的情况是对称随机徘徊(或扩散): 如果有很多质点被观察,则在充分长的时间以后,它们之中大约有一半在原点的右边,其它的则在原点左边. 尽管如此,由第三章第 5 节的反正

弦定律可知,在单个地考察时,多数质点的行为与此有偏离,它们在大部分时间内处于原点的同侧. 只要理解到统计平衡(或稳定状态)的概念并未谈到涉及个别质点性态的任何问题,无休止的讨论和错误结论就可以避免. 关于系统的趋向平衡有如下的

定理. 一个不可约的非周期马尔科夫链属于下列两类之一:

(a) 或者所有状态都是非常返的,或者所有状态都是零状态;在这种情况,当 $n \to \infty$ 时 $p_{jk}^{(n)} \to 0$ 对每一对 j, k 成立,并且不存在平稳分布.

(b) 所有的状态都是遍历的,即

$$\lim_{n\to\infty} p_{jk}^{(n)} = u_k > 0, \tag{6.2}$$

其中 u_k 是 E_k 的平均循环时间的倒数. 在这种情况, $\{u_k\}$ 是平稳分布,且不存在其它的平稳分布.

将以上的叙述稍加改变,可以把这个定理的含意弄得更清楚. 如果(6.2)成立,则对任一个初始分布 $\{a_k\}$, 有

$$a_k^{(n)} = \sum_j a_j p_{jk}^{(n)} \to u_k. \tag{6.3}$$

所以: 如果平稳分布存在,则它必定是唯一的,且不论初始分布怎样,在时刻 n 的分布总是趋向于它.

证. 由前面的定理可知, 如果状态是遍历的, 则(6.2)成立. 为了证明上面(b)中的断言,我们首先注意:

$$\sum u_k \leqslant 1. \tag{6.4}$$

这可以直接由如下事实导出: 对于固定的 j 与 n, 量 $p_{jk}^{(n)}(k=1, 2, \cdots)$ 之和为 1, 故对于每个N有 $u_1+u_2+\cdots+u_N \leqslant 1$. 在(3.3)中令 $n=1$ 并令 $m\to\infty$, 则左边趋向于 u_k, 而右边的和中的一般项则趋向于 $u_\nu p_{\nu k}$.

若只加任意有限项,则我们有

$$u_k \geqslant \sum_\nu u_\nu p_{\nu k}. \tag{6.5}$$

把这些不等式对所有的 k 求和,则在每边我们都得到有限量 $\sum u_k$. 这表明,(6.5)中的不等号是不可能的,故有

$$u_k = \sum u_j p_{jk}. \tag{6.6}$$

令 $v_k = u_k \cdot (\sum u_j)^{-1}$，则 $\{v_k\}$ 是一定稳分布，因此至少存在一个这种分布.

设 $\{v_k\}$ 是满足方程(6.1)的任一分布. 用 $p_{jk}^{(n)}$ 乘(6.1)并对所有的 j 求和，由归纳法可知，对于每个 n 有

$$v_r = \sum_{\nu} v_\nu p_{\nu r}^{(n)}. \tag{6.7}$$

令 $n \to \infty$ 我们得

$$v_r = (v_1 + v_2 + \cdots)u_r = u_r. \tag{6.8}$$

这就完成了断言(b)的证明. 如果状态为非常返的或零状态，而 $\{v_k\}$ 为一个平稳分布，则方程(6.7)成立且 $p_{\nu r}^{(n)} \to 0$，而这显然是不可能的. 所以平稳分布只可能在遍历的情况下存在，于是定理得证.

例. (a) **爱伦弗斯特模型**. 在例(2.f)中，关于平稳分布的条件取如下形式

$$u_k = \left(1 - \frac{k-1}{a}\right) u_{k-1} + \frac{k+1}{a} u_{k+1} \quad (k = 1, \cdots, a-1)$$

$$u_0 = \frac{u_1}{a}, \quad u_a = \frac{u_{a-1}}{a}. \tag{6.9}$$

容易证明，其解由伯努利分布给出:

$$u_k = \binom{a}{k} 2^{-a}. \tag{6.10}$$

这个结果可作如下解释：不论第一个容器中分子的初始数目是多少，在经过一段相当长的时间以后，在第一个容器中找到 k 个质点的概率近似地等于当 a 个分子被随机分布时恰有 k 个分子放入第一个容器中的概率(假定每个分子放入第一个容器的概率为1/2). 这是赋予我们的结果以物理意义的一个典型的例子.

当 a 很大时，二项分布的正态逼近表明，一旦极限分布(6.10)成立，我们一定能够在每个容器中找到大约一半的分子. 对物理学家来说，$a = 10^6$ 只是一个小的数字. 但即使当分子数目 $a =$

10^6 时，在一个容器内的分子多于 505000 个（密度起伏约为百分之一）的概率的数量级也只有 10^{-23}. 当 $a = 10^8$ 时，千分之一的密度起伏有同样大小的概率. 虽然系统也有偶然的机会到达可能性极小的状态，但它们的循环时间和接近平衡状态的循环时间比较起来要大得惊人. 物理中的不可逆性在如下事实中显现出来：当系统处于远离平衡的状态时，它朝向平衡移动比沿相反方向移动的可能性要大得多.

(b) **二重随机矩阵.** 一个非负元素构成的矩阵 P 称为二重随机矩阵，如果不仅其各行元素之和等于 1，而且各列元素之和亦等于 1. 如果链只含有有限个（比如说 a 个）状态，则方程组 (6.1) 显然有解 $v_k = 1/a$. 由此推出，如果一个不可约的非周期有限链有一个二重随机矩阵 P，则 $v_k = 1/a$（即在极限情况，所有状态都变成等可能的）. 在这种情况，不可能有非常返状态. 反之，在具有二重随机矩阵的不可约无限链中，所有状态或者是非常返的，或者为零状态. 为证明这个断言，假定 (6.2) 成立. 由于矩阵 P 是二重随机的，故对于每个固定的 k 及任意大的 N，我们有

$$1 = \sum_{j=1}^{\infty} p_{jk}^{(n)} \geqslant \sum_{j=1}^{N} p_{jk}^{(n)} \to N u_k. \tag{6.11}$$

由此显然有 $u_k = 0$，这与假设矛盾.（这个证明也适用于周期情况.）

(c) **循环事件.** 在例 (2.i) 中，我们已引入与任一循环事件 $\mathscr{E}$ 相关联的马尔科夫链，现在我们来证明，链的状态总是与 $\mathscr{E}$ 属于同一类型.

首先考虑 $\mathscr{E}$ 是非常返循环事件的情况；即假定 $f < 1$. 转移串 $E_i \to E_{i+1} \to E_{i+2} \to \cdots \to E_{i+n}$ 具有概率

$$\frac{1 - s_{i+1}}{1 - s_i} \cdot \frac{1 - s_{i+2}}{1 - s_{i+1}} \cdots \frac{1 - s_{i+n}}{1 - s_{i+n-1}}$$

$$= \frac{1 - s_{i+n}}{1 - s_i} \geqslant \frac{1 - f}{1 - s_i}. \tag{6.12}$$

由此可以看出，系统永不进入 E_0 的概率为正，故所有的状态是非

常返的．另一方面，当 $f=1$ 时，(6.12) 左边的项趋向于零，以概率 1 系统迟早要经过 E_0，故 E_0 是常返的．又因为每个状态都能由 E_0 到达，故链是不可约的．于是我们看到：如果 $\mathscr{E}$ 是非常返的，则链的所有状态也是非常返的；如果 $\mathscr{E}$ 是常返的，则链是不可约的，且所有状态都是常返的．

显然，链与 $\mathscr{E}$ 具有相同的周期．现在假定 $\mathscr{E}$ 是非周期的而且是常返的．我们必须判定是否存在一个平稳分布，即满足 (6.1) 的一个概率分布 $\{v_k\}$．在现在的情况，(6.1) 化归为

$$v_0=\sum_{i=0}^{\infty}\frac{f_{i+1}}{1-s_i}v_i,\quad v_k=\frac{1-s_k}{1-s_{k-1}}v_{k-1}. \tag{6.13}$$

这些方程有唯一解，即

$$v_k=(1-s_k)v_0=r_kv_0,\quad 此处\ r_k=\sum_{n=k+1}^{\infty}f_n. \tag{6.14}$$

为了使 $\sum v_k<\infty$，其充要条件是 $\sum r_k<\infty$．但 $\sum r_k=\sum nf_n$ 等于平均循环时间[参见第十一章 (1.8)]．这就证明：如果平均循环时间是无限的，则马尔科夫链的状态是零状态，如果 $\mathscr{E}$ 的平均循环时间有限，则它们的平均循环时间也有限．

我们曾由循环事件的性质导出了马尔科夫链的类似渐近性质．现在我们又已证明，每个循环事件可以用一个特殊的马尔科夫链来描述．因此，马尔科夫链的渐近性态、循环事件的渐近性态（即取整数值的独立随机变量的求和理论）、更新理论等三者实际上不过是同一解析背景的不同说法，因而本质上是等价的．

在结束本节时，让我们注意，决定是否存在平稳分布因而决定一个已给的不可约链是否遍历是比较容易的．在第 8 节中我们将导出一个区别非常返状态和常返零状态的类似准则．虽然这一问题在原则上可由讨论级数 $\sum p_{jk}^{(n)}$ 的敛散性来决定，但这在实际上是很困难的．

*7. 周　期　链

在上节中我们已把周期链的情况排除在外，但这仅仅是为了避免由于复杂的说明而掩盖了主要事实．在不可约周期链中，$p_{jk}^{(n)}$ 渐近性态的刻划容易由上节的定理中导出．为完备起见，我们将给出这样一个推导，但以后并不用到本节的结果．

根据第 5 节中的定理，不可约链的所有状态具有相同的周期 t．考虑周期为 t 的一个不可约链的任两个状态．因为每个状态可由每个其它状态到达，故存在正整数 a，b，使得 $p_{jk}^{(a)} > 0$ 及 $p_{kj}^{(b)} > 0$．现在 $p_{jj}^{(a+b)} \geqslant p_{jk}^{(a)} p_{kj}^{(b)}$，这表明，经 $a + b$ 步回到 E_j 是可能的，故 $a + b$ 必定能被周期 t 除尽．由此可知，如果由 E_j 经 a_1 步与 a_2 步都能到达 E_k，则 $a_2 - a_1$ 一定能被 t 除尽，因而用 t 除 a_1 与 a_2 时余数相同．

因此，对于固定的 E_j，每个状态 E_k 相应于某个余数 ν（此处 $0 \leqslant \nu < t - 1$），使得由 E_j 到 E_k 的转移所经的步数只可能为 ν，$\nu + t$，$\nu + 2t$，$\nu + 3t$，$\cdots$．选取 $j = 1$，我们可将所有状态分成 G_0，G_1，$\cdots$，G_{t-1} 等类；状态 E_k 属于 G_ν 的充要条件是：当且仅当 $a = \nu + nt$ 时 $p_{1k}^{(a)} > 0$．我们将 G_ν 按循环次序排列，使得 G_0 与 G_{t-1} 相邻．

易知，从 G_ν 中的状态出发，经一步转移后总是到达下一类 $G_{\nu+1}$（如果 $\nu = t - 1$，则 $G_{\nu+1}$ 为 G_0）；经两步转移后则将到达 $G_{\nu+2}$ 中的状态（从 G_{t-2} 到达 G_0，从 G_{t-1} 到达 G_1），等等．最后，经 t 步转移后必定回到属于同一类中的状态．这意味着在以 P^t 为转移概率矩阵的马尔科夫链中，类 G_ν 构成一个闭集．因为原来的链是不可约的，故每个状态能由每一其它状态到达．这意味着，在转移概率矩阵为 P^t 的链中，每个 G_ν 构成一个不可约闭集．于是我们有如下的

---

* 本节讨论专门的问题，初次阅读时可以略去．

定理. 在不可约周期马尔科夫链中,状态能分成 t 类 $G_0, \cdots, G_{t-1}$, 使得从 G_ν 中的一个状态出发, 经一步转移后总是到达 $G_{\nu+1}$ 中的一个状态(如果 $\nu = t-1$, 则到达 G_0 中的状态). 如果我们仅在时刻 $t, 2t, 3t, \cdots$来考虑这个链,则我们得到一个新链,其转移矩阵为 P^t. 在这个链中,每个 G_ν 构成一不可约闭集.

我们的定理含有关于 $p_{jk}^{(n)}$ 的渐近性态的完备知识. 如果所有的状态都是非常返状态或零状态, 则对于任意的 j, k 有 $p_{jk}^{(n)} \to 0$. 另外, 每个状态 E_k 具有有限的平均循环时间 μ_k. 设 E_j 属于 G_ν. 在 G_ν 上,我们有一个转移概率为 $p_{jk}^{(t)}$ 的不可约非周期马尔科夫链,故存在极限

$$\lim_{n\to\infty} p_{jk}^{(nt)} = \begin{cases} u_k, & \text{如果 } E_k \text{ 在 } G_\nu \text{ 中}, \\ 0, & \text{其它情况}, \end{cases} \tag{7.1}$$

此处 u_k 为在新链中 E_k 的平均循环时间的倒数, 新链的一步对应于原来链的 t 步. 于是

$$u_k = \frac{t}{\mu_k}. \tag{7.2}$$

利用 (3.2), 由 (7.1) 可求得

$$\lim_{n\to\infty} p_{jk}^{(nt+1)} = \begin{cases} u_k, & \text{如果 } E_k \text{ 在 } G_{\nu+1}, \text{中} \\ 0, & \text{其它情况}. \end{cases} \tag{7.3}$$

类似,如果 E_k 在 $G_{\nu+2}$ 中, 则 $p_{jk}^{(nt+2)} \to u_k$, 等等. 换句话说, 对于固定的 E_j 与 E_k, 序列 $p_{jk}^{(n)}$ 是渐近周期的; 在其中每隔 $t-1$ 个接连的零就出现一个正元素,且它们收敛于 $u_k = 1/\mu_k$.

由第 6 节中的定理可知, 在每个类 G_ν 中, u_k 之和为 1. 因为有 t 个类. 故由 (7.2) 可知, 序列 $\{1/\mu_k\}$ 表示一个概率分布. 第 6 节中的论证直接表明,这个分布是平稳的,且不存在其它的平稳分布.

8. 非常返状态

从常返状态 E_j 出发, 系统只可能到达 E_j 的闭包中的常返状

态 E_k,而且在这种情况下,我们已经得到关于 $p_{jk}^{(n)}$ 的完备知识.

如果 E_j 是非常返的,而 E_k 是遍历的,则由 (5.7) 有

$$p_{jk}^{(n)} \to \mu_k^{-1} f_{jk}, \tag{8.1}$$

其中 μ_k 为 E_k 的平均循环时间, f_{jk} 为从 E_j 出发,系统迟早将到达 E_k 的概率. 然而, E_k 属于一个不可约子链 C, 并且从 E_k 出发,系统必定到达 C 中的每个状态. 所以,对于每个固定的 j, 概率 f_{jk} 对于 C 中所有状态来说都是一样的. 换句话说, 如果 C 是一个具有遍历状态的不可约子链, 而 E_j 是非常返的, 则对于 C 中的每个 E_k, 有

$$p_{jk}^{(n)} \to \mu_k^{-1} x_j, \tag{8.2}$$

其中 x_j 为从 E_j 出发,系统迟早将进入 C 的概率. 显然,对零状态来说,(8.2) 的右边必须用 0 来代替,且在 E_k 具有周期的情况仅需作通常的例行修改.

为了完全描绘出马尔科夫链的渐近状态,还须解决如下的

三个问题. (a) 已给一非常返状态 E_j 及一常返闭集 C, 试求从 E_j 出发, 系统迟早将进入 C(即到达 C 中的一个状态)的概率 x_j.

(b) 求出系统将永远停留在非常返状态集合中的概率 y_j.

(c) 已给一不可约链,确定其状态是常返的还是非常返的.

我们将看到,稍加改动后, 问题 (c) 变成问题 (a) 的一个特殊情况.

设 T 为所有非常返状态的集合,并假定,在开始时系统处于非常返状态 E_j; 设 $x_j^{(n)}$ 为系统在时刻 n 第一次到达闭集 C 的概率. 则

$$x_j = \sum_{n=1}^{\infty} x_j^{(n)} \tag{8.3}$$

为系统迟早将到达 C 并停留在 C 中的概率. 借用简单的随机徘徊的术语,我们将称 x_j 为被 C 吸收的概率. 差 $1-x_j$ 说明被其它的闭集吸收与(在某些无限链的情况)永远停留在非常返状态集合中的可能性.

显然,

$$x_j^{(1)} = \sum_C p_{jk}, \tag{8.4}$$

这里的和是对所有在 C 中的状态 E_k 的下标 k 来求的. 如果系统在第 $n+1$ 步到达 C 中,则第一步必须由 E_j 到达另一非常返状态. 因此显然有

$$x_j^{(n+1)} = \sum_T p_{j\nu} x_\nu^{(n)}, \tag{8.5}$$

现在的和是对有非常返状态 E_ν 的下标 ν 来求的. 方程 (8.4) 与 (8.5) 构成递推关系, 由它们可以唯一地确定 $x_j^{(n)}$. 将 (8.5) 对一切 $n=1, 2, 3, \cdots$ 求和, 即可看出, 吸收概率 x_j 是线性方程组

$$x_j - \sum_T p_{j\nu} x_\nu = x_j^{(1)} \tag{8.6}$$

的解.

于是我们得到问题 (a) 的一个回答; 概率 x_j 可以构造性地由 (8.3)—(8.5) 来确定; 但把它们看作是线性方程组 (8.6) 的解更好. 在这种关系下,唯一性问题就被提出,但幸而它是问题 (b) 的特殊情况.

设 $y_j^{(n)}$ 为系统在时刻 n 处于非常返状态的概率. 显然

$$y_j^{(1)} = \sum_T p_{j\nu},$$
$$y_j^{(n+1)} = \sum_T p_{j\nu} y_\nu^{(n)}, \tag{8.7}$$

此处的和也是对所有非常返状态 E_ν 的下标 ν 来求的. 由 (8.7) 可以推出, $y_j^{(1)} \leqslant 1$. 因而 $y_j^{(2)} \leqslant y_j^{(1)}$, 一般地 $y_j^{(n+1)} \leqslant y_j^{(n)}$. 所以极限

$$y_j = \lim_{n\to\infty} y_j^{(n)} \tag{8.8}$$

存在; y_j 为系统永远停留在非常返状态中的概率. 由 (8.7) 我们有

$$y_j = \sum_T p_{j\nu} y_\nu. \tag{8.9}$$

因此概率 y_i 满足方程(8.9)，但这并未解决主要问题，即是否对所有的 i 有 $y_i=0$. 设方程(8.9)存在一个有界解，比方说，

$$z_j=\sum_T p_{j\nu}z_\nu, \quad |z_\nu|\leqslant 1. \tag{8.10}$$

比较(8.10)和(8.7)可得 $|z_i|\leqslant y_i^{(1)}$，因此由归纳法可知，$|z_i|\leqslant y_i^{(n)}$ 对一切 n 成立. 由此推出，$y_i=0$ 对一切 i 成立的充分必要条件是，方程组(8.10)无非零解. 最后，如果线性方程组(8.6)有两组不同的解，则它们的差就会是线性方程组(8.10)的一组非零解. 于是我们有

定理 1. 问题(a)的概率是线性方程组(8.6)的解. 如果从任何非常返状态 E_j 出发，系统永远停留在非常返状态的概率 y_i 为 0，则这个解是唯一的. y_i 满足方程(8.9).

注. 我们已经看到概率 y_i 可以看作是(8.9)的以 1 为界的解；$\{x_i\}$ 具有类似的性质.

推论. 在有限马尔科夫链中，系统永远停留在非常返状态的概率为零. 从非常返状态出发，进入一个闭集 C 的概率 x_i 是线性方程组(8.6)的唯一解.

证. 我们必须证明方程(8.9)无解. 用反证法，设 M 是有限多个 y_i 中的最大者. 不失一般性，可将各状态适当排列，使得 y_i 按递减的次序出现，比如说 $y_1=y_2=\cdots=y_a=M>y_{a+1}\geqslant y_{a+2}\geqslant\cdots$. 由(8.9)可知，如果 $i\leqslant a$，则我们有

$$M=\sum_T p_{i\nu}y_\nu=\sum_{\nu=1}^{a} p_{i\nu}M+\sum_{\nu\geqslant a+1} p_{i\nu}y_\nu. \tag{8.11}$$

这个等式只可能在对一切 $\nu>a$ 都有 $p_{i\nu}=0$ 时成立. 在这种情况，$E_1,\cdots,E_a$ 构成一个闭集，而这是不可能的，因为一个有限链必定包含常返状态(第 5 节的推论).

定理 1 可用来计算吸收概率，即进入一个吸收状态的概率.

例. (a) **具有吸收壁的随机徘徊**[例(2.b)]. 取吸收状态 E_0 为 C. 则 $x_1^{(1)}=q$ 且当 $j>1$ 时 $x_j^{(1)}=0$. 方程组(8.6)变成

$$x_1-px_2=q,$$

$$x_j - qx_{j-1} - px_{j+1} = 0 \quad (j = 2, 3, \cdots, a-2),$$
$$x_{a-1} - qx_a = 0. \tag{8.12}$$

这与第十四章中的方程组(2.1)—(2.2)是一样的，故其解由第十四章中的(2.4)给出.

(b) **序贯抽样**[例(2.j)]. 再次设 C 为状态 E_0. 则 $x_j^{(1)} = r_j$，于是方程(8.6)化归为第十四章中的(8.2)(其中 u_x 代表现在的 x_j；也可参看第十四章中的问题 4).

(c) **遗传学**[例(2.k)]. 这里两个状态 E_0 与 E_{2N} 中的每一个都构成闭集. 在 E_0 与 E_{2N} 处的吸收分别意味着总体最终仅由 aa 型或仅由 AA 型个体组成. 对于在 E_0 的吸收，我们有 $x_j^{(1)} = p_{j0} = (1 - j/2N)^{2N}$，于是(8.6)取如下形式

$$x_j - \sum_{\nu=1}^{2N-1} \binom{2N}{\nu} \left(\frac{j}{2N}\right)^{\nu} \left(1 - \frac{j}{2N}\right)^{2N-\nu} x_\nu = \left(1 - \frac{j}{2N}\right)^{2N}. \tag{8.13}$$

似乎有理由认为,如果 A 型基因与 a 型基因的比例为 $j:2N-j$，则它们的成活机会也具有相同的比例. 如果这是对的,则(8.13)的解必须为 $x_j = 1 - j(2N)^{-1}$. 在(8.13)中认出均值为 j 的二项分布的项后就容易证明,这些 x_j 确实满足(8.13).

最后,我们给出问题(c)的一个解答.

定理 2. 设一个不可约链有状态 $E_0, E_1, \cdots$. 为了使状态是非常返的,其必要充分条件是方程组

$$y_i = \sum_{j=1}^{\infty} p_{ij} y_j, \quad i = 1, 2, \cdots \tag{8.14}$$

有一个非零有界解.

证. 在 $\{y_i\}$ 的构造式(8.7)—(8.8)中，将 T 解释为状态 $E_1, E_2, \cdots$ 的集合(E_0 的补集). 前面的证明无需改变就可以适用于现在的情况，于是我们看到，停留在 T 中(即不进入 E_0)的概率由(8.8)给出,且满足(8.9).

例. (d) **自由随机徘徊.** 例(2.e)需要作一些记号上的改变，因为状态是由 $-\infty$ 到 $+\infty$ 来编号的. 不过,很清楚,我们的准则

依赖于方程

$$y_i = py_{i+1} + qy_{i-1}, \quad y_0 = 0, \quad i = \pm 1, \ \pm 2, \ \cdots \tag{8.15}$$

的解的唯一性. 显然,所有的 y_i 能由 y_1 与 y_{-1} 递推地计算. 如果 $p > q$, 则

$$y_i = \left\{1 - \left(\frac{q}{p}\right)^i\right\} y_1, \quad y_{-i} = 0, \quad i = 1, 2, \cdots \tag{8.16}$$

是唯一解且是有界的. 如果 $p = q$, 则解为 $y_i = iy_1$ 且是无界的. 这里我们用马尔科夫链的方法导出了旧的结果: 如果 $p \neq q$, 则状态是非常返的;如果 $p = q$, 则状态是常返的.

(e) 考虑矩阵

$$P = \begin{bmatrix} q_0 & p_0 & 0 & 0 & 0 & \cdots \\ q_1 & 0 & p_1 & 0 & 0 & \cdots \\ 0 & q_2 & 0 & p_2 & 0 & \cdots \\ 0 & 0 & q_3 & 0 & p_3 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \end{bmatrix}.$$

它表示 $(0, \infty)$ 中具有变转移概率的一个随机徘徊. 它在第十七章所讨论的生灭过程的理论中起着重要作用. 方程(8.14)化归为

$$y_1 = p_1 y_2, \quad y_i = q_i y_{i-1} + p_i y_{i+1}, \quad i = 2, 3, \cdots \tag{8.17}$$

并且可以逐步地解出,因为

$$\frac{y_{i+1} - y_i}{y_i - y_{i-1}} = \frac{q_i}{p_i}, \tag{8.18}$$

所以

$$y_{i+1} - y_i = y_1 \cdot \frac{q_1}{p_1} \cdot \frac{q_2}{p_2} \cdots \frac{q_i}{p_i}. \tag{8.19}$$

把这些等式相加,我们看到,存在一个有界解的充要条件是 $\sum L_i < \infty$, 其中 $L_i = (q_1 \cdots q_i)(p_1 \cdots p_i)^{-1}$. 所以, 如果 $\sum L_i < \infty$, 则状态为非常返的,否则为常返的.

9. 在洗牌问题中的应用

编号为 $1, 2, \cdots, N$ 的 N 张牌能排成 $N!$ 种不同的次序，每种次序表示系统的一种可能状态．每次洗牌使现存的状态转移到某个其它状态．例如，“错牌”将把次序 $(1, 2, \cdots, N)$ 变成循环等价次序 $(r, r+1, \cdots, N, 1, 2, \cdots, r-1)$ 之一．而逆序经错牌后则变成 $(N-r+1, N-r, \cdots, 1, N, N-1, \cdots, N-r+2)$．换句话说，我们把洗牌看成是转移 $E_j \to E_k$．如果同种方式的洗牌重复进行，则系统（从给定的状态 E_j 出发）将顺序地经过一列状态，并且在有限步以后，原来的次序将被重新建立．从那时以后，同一列接连的状态将周期地重复出现．对大多数洗牌方式来说，周期是相当小的，并且决不是所有状态都能由这种程序所达到[1]．例如，一次精确的“插牌”把 $2m$ 张牌从 $(1, \cdots, 2m)$ 变成 $(1, m+1, 2, m+2, \cdots, m, 2m)$．对于六张牌来说，四次精确的插牌就将重建原来的次序．对于十张牌来说，原来的次序将在六次精确的插牌后出现，所以，对十张牌重复进行精确的插牌仅能得到 $10! = 3628800$ 种可能次序中的六种．

实际上，玩牌人或许希望改变洗牌方式，并且，无论如何，偶然变动总是由机会引入．我们将假定，我们能够通过假定每种特别的洗牌方式有某个概率（可能为零）来说明玩牌人的习惯和机会的影响．关于这些概率的数值我们无需作任何假定，但假定玩牌人洗牌时不考虑过去的情况，也不知道牌的次序[2]．这意味着，接连的洗牌对应于具有固定概率的独立试验；于是，对于实际的一副牌来说，我们有一个马尔科夫链．

现在我们证明，转移概率矩阵 P 是二重随机矩阵［例 (6.b)］．

1) 用群论的语言来说，这相当于说置换群不是循环的，因而不能由一个简单运算产生．

2) 这个假设对应于桥牌中的相应情况．容易设计出更复杂的洗牌技术，在其中，每次洗牌的方式依赖于以前的洗牌方式，在这种情况，最后的结果不是马尔科夫链［参见例 (10.e)］．

事实上,如果一次洗牌变状态(牌的次序) E_j 为E_k, 则存在洗牌使之变成 E_j 的、另一状态 E_r. 这意味着, 除了次序不同外, P 的第 j 列与第 j 行是恒等的. 故所有列和为 1.

由此可知, 不可能有非常返状态存在. 如果链是不可约和非周期的,则在极限情况,所有的状态变成等可能的. 换句话说,任一种洗牌方式,只要它产生一个不可约的非周期链,都是合适的. 可以假定,通常情况就是这样. 不过,设一副牌共有偶数张,并且洗牌的方法是,将它们等分为两组,并用任意的方法分别洗这两组牌. 如果将这两组牌按它们原来的次序拼在一起, 则马尔科夫链是可约的(因为不是每个状态都能由每个其它状态到达). 如果两部分的次序被颠倒,则链有周期 2.

我们看到,有理由指望, 连续不断的洗牌会产生完全的"随机性"并消除原始次序的痕迹. 然而, 应当指出, 为此所需的洗牌次数是极大的[34].

10. 一般的马尔科夫过程

在应用中用随机变量的术语来描述马尔科夫链通常是方便的. 简单地用整数 k 来代替上节中的符号 E_k 就能做到这一点. 于是系统在时刻 n 的状态是以概率 $a_k^{(n)}$ 取值 k 的一个随机变量 $\mathbf{X}^{(n)}$; $\mathbf{X}^{(n)}$ 与 $\mathbf{X}^{(n+1)}$ 的联合分布由

$$\mathbf{P}\{\mathbf{X}^{(n)}=j,\ \mathbf{X}^{(n+1)}=k\}=a_j^{(n)}p_{jk}$$

给出,而 $(\mathbf{X}^{(0)},\cdots,\mathbf{X}^{(n)})$ 的联合分布则由(1.1)给出. 指定 E_k 以一个异于 k 的数值也是可能的, 而且有时更好一些. 根据这种表示法,马尔科夫链变成一个特殊的随机过程,即(相依)随机变量的一个序列[1] $(\mathbf{X}^{(0)},\mathbf{X}^{(1)},\cdots)$. 上标 n 起着时间的作用. 在第十七章中我们对时间参数连续变化的、更一般的随机过程作初步的介绍. "马尔科夫过程"这一术语适用于很大而且很重要的一类随

1) 明确的表达要参照无穷乘积空间, 但实际上我们仅涉及有限多个随机变量的联合分布.

机过程(时间参数可以是离散的，也可以是连续的). 甚至在离散的情况，也存在比到现在为止我们所研究过的简单链更一般的马尔科夫过程. 因此,给出马尔科夫性质的定义,指出刻划马尔科夫链的特殊条件，最后，给出几个非马尔科夫过程的例子,这些都是有用的.

在概念上,马尔科夫过程是经典力学的过程的概率模拟,在其中将来的发展完全决定于现在的状态而与现在的状态是怎样发展来的无关.力学过程与可塑性等理论中出现的有后效过程(或遗传过程)形成鲜明的对照，在后者中，系统整个过去的历史影响它的将来. 在随机过程中,将来的情况决不能唯一地确定,但至少有概率关系使我们能作出预言. 对于本章所研究的马尔科夫链来说，有关将来的概率关系显然仅依赖于现在的状态，而与现在的状态是怎样从过去变来的无关. 换句话说，如果两独立的系统具有相同的转移概率且恰好处于同一状态，则与它们将来的发展有关的所有概率都是恒等的. 以上的描述相当含糊，现在我们给出其精确定义如下:

定义. 一列取离散值的随机变量称为一个马尔科夫过程，如果对于每个整数 $n_1 < n_2 < \cdots < n_r < n$ 的有限集合，$(\mathbf{X}^{(n_1)}, \mathbf{X}^{(n_2)}, \cdots, \mathbf{X}^{(n_r)}, \mathbf{X}^{(n)})$ 的联合分布用如下的方式来定义：在假设 $\mathbf{X}^{(n_1)} = x_1, \cdots, \mathbf{X}^{(n_r)} = x_r$ 下,关系 $\mathbf{X}^{(n)} = x$ 的条件概率恒等于在单一假设 $\mathbf{X}^{(n_r)} = x_r$ 下 $\mathbf{X}^{(n)} = x$ 的条件概率. 这里，$x_1, \cdots, x_r$ 是使假设具有正概率的任意数.

这个定义可用较简洁的语言表述如下：设在时刻 n_r 已给状态 x_r，则关于系统以前的状态的任何附加知识都不能改变在将来的一个时刻 n 的状态 x 的(条件)概率.

本章所研究的马尔科夫链显然是马尔科夫过程，但它们具有不包含在定义中的如下附加性质. 对于上节所研究的马尔科夫链来说,转移概率 $p_{jk} = \mathbf{P}\{\mathbf{X}^{(m+1)} = k \mid \mathbf{X}^{(m)} = j\}$ 不依赖于 m. 更一般的转移概率

$$p_{jk}^{(n-m)} = \mathbf{P}\{\mathbf{X}^{(n)} = k \mid \mathbf{X}^{(m)} = j\} \quad (m < n) \tag{10.1}$$

则仅依赖于差 $n-m$. 在这种情况,我们说转移概率是平稳的(或不变的). 对于一般的整数值马尔科夫链来说,(10.1)的右边依赖于 m 与 n. 我们将用 $p_{jk}(m,n)$ 来表示它,所以 $p_{jk}(n,n+1)$ 是在时刻 n 的一步转移概率. 代替(1.1),我们得到路径 $(j_0,j_1,\cdots,j_n)$ 的概率的如下表达式:

$$a_{j_0}^{(0)}p_{j_0j_1}(0,1)p_{j_1j_2}(1,2)\cdots p_{j_{n-1}j_n}(n-1,n). \tag{10.2}$$

显然(3.3)的适当推广是如下等式

$$p_{jk}(m,n)=\sum_{\nu}p_{j\nu}(m,r)p_{\nu k}(r,n), \tag{10.3}$$

它对于满足不等式 $m<r<n$ 的所有 r 成立. 这个恒等式可直接由马尔科夫过程的定义也可由(10.2)得到;它称为**查普曼-柯尔莫果洛夫方程.**

在本章中我们基本上解决了高阶转移概率的渐近性态问题,但只有少数几个已建立的性质对于一般的离散马尔科夫过程成立. 所以我们对一般理论不作详细讨论.

非马尔科夫过程的例子.

(a) **波利亚罐子模型**[第五章例(2.c)]. 设 $\mathbf{X}^{(n)}$ 等于1或0按照第 n 次抽出的球是黑球还是红球而定. 序列 $\{\mathbf{X}^{(n)}\}$ 不是马尔科夫过程. 例如

$$\mathbf{P}\{\mathbf{X}^{(3)}=1\,|\,\mathbf{X}^{(2)}=1\}=(b+c)/(b+r+c),$$

但

$$\mathbf{P}\{\mathbf{X}^{(3)}=1\,|\,\mathbf{X}^{(2)}=1,\mathbf{X}^{(1)}=1\}=(b+2c)/(b+r+2c).$$

(参看第五章问题19—20). 另一方面,如果 $\mathbf{Y}^{(n)}$ 是在时刻 n 罐中的黑球数,则 $\{\mathbf{Y}^{(n)}\}$ 是具有常转移概率的通常马尔科夫链.

(b) **高阶和.** 设 $\mathbf{Y}_0,\mathbf{Y}_1,\cdots$ 是相互独立的随机变量,令 $\mathbf{S}_n=\mathbf{Y}_0+\cdots+\mathbf{Y}_n$. 则差 $\mathbf{S}_n-\mathbf{S}_m$ 仅依赖于 $\mathbf{Y}_{m+1},\cdots,\mathbf{Y}_n$,并且容易看出,序列 $\{\mathbf{S}_n\}$ 是一马尔科夫过程. 现在我们进一步定义一个新的随机变量序列 $\mathbf{U}_n$ 为 $\mathbf{U}_n=\mathbf{S}_0+\mathbf{S}_1+\cdots+\mathbf{S}_n$(这意味着 $\mathbf{U}_n=\mathbf{Y}_n+2\mathbf{Y}_{n-1}+3\mathbf{Y}_{n-2}+\cdots$). 序列 $\{\mathbf{U}_n\}$ 构成一个随机过程,其概率关系原则上能用 $\mathbf{Y}_k$ 的分布来表示. 过程 $\{\mathbf{U}_n\}$ 通常不

是马尔科夫型的,因为,例如,没有理由断定 $\mathbf{P}\{\mathbf{U}_n = 0 | \mathbf{U}_{n-1} = a\}$ 与 $\mathbf{P}\{\mathbf{U}_n = 0 | \mathbf{U}_{n-1} = a, \mathbf{U}_{n-2} = b\}$ 相同;根据关于 $\mathbf{U}_{n-1}$ 与 $\mathbf{U}_n$ 的知识比单独根据关于 $\mathbf{U}_{n-1}$ 的知识能作出较好的预言.

在时间参数连续的情况，前面的求和用积分来代替. 在扩散理论中, $\mathbf{Y}_n$ 起着加速度的作用; $\mathbf{S}_n$ 则为速度,于是 $\mathbf{U}_n$ 为位置. 如果仅仅位置能够测量,则我们只好研究一个非马尔科夫过程,虽然它是间接地用一个马尔科夫过程来定义的.

(c) **移动平均**. 再次令 $\{\mathbf{Y}_n\}$ 表示相互独立的随机变量序列. r 阶移动平均定义为 $\mathbf{X}^{(n)} = (\mathbf{Y}_n + \mathbf{Y}_{n+1} + \cdots + \mathbf{Y}_{n+r-1})/r$. 容易看出, $\mathbf{X}^{(n)}$ 不是马尔科夫过程. 在很多应用中都遇到这种类型的过程(参见问题 26).

(d) **交通问题**. 为了得到一个非马尔科夫过程的经验例子, R. 弗思[35] 对某街道上的行人进行了广泛的观察. 这个过程的一个理想化的数学模型能用如下的方法得到. 为简单起见，我们假定所有的行人有同一速度 v; 又我们仅考虑沿一个方向走动的行人. 在时刻 $t = 0$, 我们将正 x 轴分为长为 δ 的区间,每个区间可能包含也可能不包含一个行人. 我们假定，在各区间中的行人的分布由一列伯努利试验来确定. 换句话说，我们有独立随机变量 $\mathbf{Y}_k$ 的一个序列，其中每一个分别以概率 p 与 q 取值 1 或 0. 如果 $\mathbf{Y}_k = 1$, 则区间 $(k-1)\delta \leqslant x \leqslant k\delta$ 包含一个行人. 现在设整个轴以速度 v 沿负向运动，让我们观察长为 $N\delta$ 的固定区间中的行人数，这个区间在时刻 $t = 0$ 时被运动的 x 轴的区间 $0 \leqslant x \leqslant N\delta$ 所覆盖. 在时刻 t 这个固定区间被 x 轴的区间 $vt \leqslant x \leqslant vt + N\delta$ 所覆盖. 设在时刻 $n\delta/v$ 进行观察，并设 $\mathbf{X}^{(n)}$ 为在时刻 n 在我们的固定区间中被观察到的行人数. 则 $\mathbf{X}^{(n)} = \mathbf{Y}_n + \mathbf{Y}_{n+1} + \cdots + \mathbf{Y}_{n+N-1}$, 所以我们的过程，除因子 $1/N$ 外，是一个移动平均过程. 因此它不是马尔科夫过程. (当 $\delta \to 0$ 时,我们得到一个连续模型,在其中普阿松分布取代二项分布而起作用.)

(e) **马尔科夫过程的迭加**(合成洗牌).

有很多专门装置(例如电话局、计数器、滤波器中的选择器

群)，它们的动作可用两个马尔科夫过程的迭加(其结果是一非马尔科夫过程)来描述. 从如下洗牌方法的研究中可以得到关于这种机构的清楚了解.

设有甲、乙两组同样的牌,每组各有N张. 洗牌通常是对乙组牌施行. 如果它的各张牌的次序是 $(a_1, a_2, \cdots, a_N)$，则我们将甲组牌加以排列，使得第一、第二、…、第N张牌分别调动到第 $a_1, a_2, \cdots, a_N$ 个位置.于是洗乙组牌就间接地决定了甲组牌的次序. 后者构成一个非马尔科夫型的随机过程. 为证明这一点，只需证明,关于甲组牌的两个接连的次序的知识对未来所提供的线索,一般要比关于最后的次序的单一知识所能提供的要多. 我们就一个简单的特殊情况来证明这一点.

设$N=4$，且假定在初始时刻乙组牌的次序为(2431). 此外，假定采用"错牌"的方法来洗牌,即次序 (a_1, a_2, a_3, a_4) 变成 (a_2, a_3, a_4, a_1)，(a_3, a_4, a_1, a_2)，(a_4, a_1, a_2, a_3) 这三个次序之一;我们赋予这三种可能中的每一种以概率 1/3. 由于这些约定，在任何时刻，乙组牌的次序总是(2431)，(4312)，(3124)，(1243)这四种次序之一. 另一方面,简单的实验表明,甲组牌将逐步经过所有 24 种可能次序,且它们之中的每一个将结合乙组牌的四个可能次序中的每一个出现. 这意味着,甲组牌的次序(1234)将极为经常地重复出现,且在它之后总是出现(2431),(4312),(3124),(1243)这四个次序之一. 现在乙组牌永远不能保持同一次序，故甲组牌不能接连两次遭受同一种排列. 因此,如果在时刻$(n-1)$与n，甲组牌的次序分别为(1234)与(1243)，则在时刻$n+1$状态(1234)是不可能的. 这样,关于在时刻$(n-1)$与n的状态的知识所传达的信息，要比关于在时刻n的状态的单一知识所传达的要多.

*11. 杂　录

(a) 逆概率. 虽然,最自然的是研究一个系统的将来发展,但

* 初次阅读时本节可以略去.

有时也必需研究它的过去. 考虑具有状态 E_k 及不变转移概率 p_{jk} 的一个马尔科夫链, 它在时刻 n 的绝对概率为 $a_k^{(n)} = \sum a_v^{(0)} p_{vk}^{(n)}$. 在已知系统在时刻 n 处于状态 E_k 的条件下, 系统在时刻 $m < n$ 处于状态 E_j 的条件概率为(与时刻 n 以后的状态无关)

$$q_{kj}(n, m) = \frac{a_j^{(m)}}{a_k^{(n)}} p_{jk}^{(n-m)}, \quad m < n. \tag{11.1}$$

这个公式仅当 $a_k^{(n)} > 0$ 时才有意义; 否则问题中的条件概率没有定义. 如果所有的 $a_k^{(n)}$ 都是正的, 则 (11.1) 定义一组具有马尔科夫过程所需的一切性质的转移概率. 特别, $q_{kj}(n, m)$ 满足时间反向的查普曼-柯尔莫果洛夫等式 (10.3), 即

$$q_{kj}(n, m) = \sum_v q_{kv}(n, r) q_{vj}(r, m) \tag{11.2}$$

$(m < r < n)$. $q_{kj}(n, m)$ 称为逆概率[36]. 特别, 考虑一个具有平稳概率 $\{u_k\}$ 的不可约链. 则对于所有的 n, $a_k^{(n)} = u_k$, 且 $u_k > 0$ (参看第 6 节与第 7 节). 在这种情况下, 一步转移 $q_{k,j}(n+1, n)$ 不依赖于 n 且可以简化为

$$q_{kj} = \frac{u_j}{u_k} p_{jk}. \tag{11.3}$$

$\{q_{kj}\}$ 是一随机矩阵, 所以这里的逆概率定义一个具有不变转移概率的马尔科夫链. 如果 $q_{kj} = p_{jk}$, 则原来的链称为可逆的; 它的概率关系在时间上是对称的.

(b) 中心极限定理. 循环事件的理论包含关于马尔科夫链的进一步的知识. 设 E_k 为一个固定的常返状态, 其循环时间具有有限的方差 σ_k^2 (如果链是有限的, 则这个条件总成立). 设 $\mathbf{N}_n$ 为直到时刻 n 为止系统经过 E_k 的次数. 则由第八章第 6 节可知, 变量 $\mathbf{N}_n$ 是渐近正态分布的. 根据本章的记号, 我们有 $\mathbf{E}(\mathbf{N}_n) = 1/\mu_k = u_k$; 在有限链的情况计算方差的方法将在下章中给出. 特别, 当 $n \to \infty$ 时, 对于任意的 $\varepsilon > 0$, $\left|\frac{\mathbf{N}_n}{n} - u_k\right| < \varepsilon$ 的概率趋向

于1. 这是关于经过 E_k 的次数的弱大数定律. 类似,强大数定律与迭对数定律也成立, 且无需特别的证明. 在无限链的情况, E_k 的循环时间即使有有限的平均值,也未必有有限的方差. 然而,关于循环事件的一般极限定理在这种情况下适用.

随机变量 $\mathbf{N}_n$ 可以定义为 $\mathbf{N}_n = \mathbf{X}_1 + \cdots + \mathbf{X}_n$, 其中当系统在时刻 n 处于状态 E_k 时, $\mathbf{X}_n$ 等于 1, 否则 $\mathbf{X}_n$ 等于零. 这种表示有如下推广. 我们赋予状态 E_k 以任意的数值 x_k, 并且当系统在时刻 n 处于状态 E_k 时令随机变量 $\mathbf{X}_n$ 等于 x_k. 和通常一样,我们令 $\mathbf{S}_n = \mathbf{X}_1 + \cdots + \mathbf{X}_n$. 对于有限马尔科夫链, 杜勒林[37] (W. Doeblin) 曾证明, 一般地说, 中心极限定理和迭对数定律对于 $\mathbf{S}_n$ 成立. 仅当数 x_ν 选择得使对于每个由 E_k 返回 E_k 的最短路径, x_ν 之和等于一个与路径无关的常数 c 时,出现一个例外.

(c) **非随机矩阵**. 本章的定理描述任意的随机矩阵 P [即其元素满足条件 (1.2) 的矩阵]的幂 P^n 的渐近性态. 容易将这些结果推广到一类更一般的矩阵. 设 P 是任一个具有非负元素的(有限或无限)矩阵,并用 S_j 表示其行和, 即 $S_j = \sum_k p_{jk}$. 我们假定序列 S_j 有界,即存在一个常数 M, 使得 $S_j \leqslant M$. 在这些条件下, P^n 的渐近性态仍然由我们的定理来描述, 这是因为 P 能化归为随机矩阵.

为确定起见,设 P 的行与列从 1 开始编号,并且首先考虑对于所有的 j, $S_j \leqslant 1$ 的情况. 在这种情况,我们在 P 中加上标号为 0 的一行和一列而将 P 扩大, 所增加的元素为 $p_{00} = 1$, $p_{01} = p_{02} = \cdots = 0$, 且当 $j \geqslant 1$ 时 $p_{j0} = 1 - S_j$. 扩大后的新矩阵 Q 是一随机矩阵, 故它们渐近性态由我们的定理给出. 另一方面, P^n 是 Q^n 的角元素 $p_{00}^{(n)}$ 的子矩阵. 在一般情况,行和 S_j 可能大于 1, 但这时我们可以用元素为 p_{jk}/M 的矩阵 P^* 来代替矩阵 P. P^* 的行和 S_j^* 满足条件 $S_j^* \leqslant 1$, 于是我们能够描述幂 P^{*n} 的渐近性态. 然而, 矩阵 P^n 与 P^{*n} 仅相差一个因子 M^n, 所以我们的定理实际上在所有的情况都描述了 $p_{jk}^{(n)}$ 的渐近性态.

在广义随机徘徊中遇到上述类型的矩阵.

(**d**) **文献**. 关于有限马尔科夫链有大量的文献. 在 M. 弗雷谢[1] 的综合性论文中可以找到各种着手方法的详细说明和早期工作的文献. 有限链的代数处理将在下章中给出. 有限链的整个理论可由弗洛必留斯 (Frobenius) 的具有正元素的矩阵理论中导出. 这个方法特别被罗曼洛夫斯基 (В. И. Романовский) 所开拓. 遗憾的是,这些方法不能应用于首先被柯尔莫果洛夫[2]所考虑的无限链的更有趣的情况. 他的工作被 W. 杜勃林[41]与 J. L. 杜勃[42]所继续. 钟开莱[43] 特别研究了当某些状态被禁止时, 由一个状态到另一个状态的转移. 德曼[44] (C. Dermán) 证明,在具有零状态的不可约链中,关于平稳分布的方程 (6.1) 容有使得 $\sum v_k = \infty$ 的唯一解 $\{v_k\}$. 反向公式对这样的解也有意义, 并且近代理论日益关心类似的无界解的使用[45]. T. E. 哈里斯[46] (Harris) 系统地研究了与马尔科夫链有关的几类很一般的非马尔科夫过程.

12. 问　　题

1. 在伯努利试验序列中, 如果第 $n-1$ 次与第 n 次试验的结果为 SS, 则我们说在时刻 n 观察到状态 E_1. 类似, E_2, E_3, E_4 分别代表 SF, FS, FF. 试求矩阵 P 及它的所有幂. 推广这个概型.

2. 将下列四个链的状态加以分类, 这些链的矩阵 P 的各行分别给出如下. 在每种情况下,求 P^2 及 $p_{jk}^{(n)}$ 的渐近性态.

(a) $(0, \frac{1}{2}, \frac{1}{2})$, $(\frac{1}{2}, 0, \frac{1}{2})$, $(\frac{1}{2}, \frac{1}{2}, 0)$;

(b) $(0, 0, 0, 1)$, $(0, 0, 0, 1)$, $(\frac{1}{2}, \frac{1}{2}, 0, 0)$, $(0, 0, 1, 0)$;

(c) $(\frac{1}{2}, 0, \frac{1}{2}, 0, 0)$, $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}, 0, 0)$, $(\frac{1}{2}, 0, \frac{1}{2}, 0, 0)$, $(0, 0, 0, \frac{1}{2}, \frac{1}{2})$, $(0, 0, 0, \frac{1}{2}, \frac{1}{2})$;

(d) $(0, \frac{1}{2}, \frac{1}{2}, 0, 0, 0)$, $(0, 0, 0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, $(0, 0, 0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, $(1, 0, 0, 0, 0, 0)$, $(1, 0, 0, 0, 0, 0)$, $(1, 0, 0, 0, 0, 0)$.

3. 在掷一颗均匀的骰子时,如果 j 是前 n 次投掷中出现的最大点数,则

1) 见 M. Fréchet [38]. 关于马尔科夫链的另一专著见 B. Hostinsky [39].
2) 见 A. Kolmogorov [40].

我们说在时刻 n 系统处于状态 E_j. 试求矩阵 P^n 并证明公式 (3.3) 成立。

4. 在例 (2.1) 中，试求从 E_k 出发，系统分别在 E_1 与 E_5 结束的(吸收)概率 x_k 与 $y_k(k=2,3,4,6)$.（直接从基本定义来解这个问题，而无需参考第 8 节的公式。）

5. 利用马尔科夫链来讨论第一章中的例 (5.b). 计算每个赌徒获胜的概率。

6. 设 P 的第一行为 $\{p_1, p_2, \cdots\}$. 在以下各行中，$p_{j,j-1}=1$，其余各元素均为零。讨论各状态的特征；如果平稳分布存在，试求之。

7. 设 P 的第一列为 $\{q_0, q_1, \cdots\}$ 且 $p_{i,i+1}=1-q_i(i=0,1,\cdots)$. 试证明当且仅当 $\sum q_i<\infty$ 时，状态是非常返的. 何时状态为零状态? 如果存在平稳分布，试求之。

8. **一个反射壁**. 考虑如下的随机徘徊矩阵，其中 $p_{k,k+1}=p, p_{k,k-1}=q$ $(k=2,3,\cdots)$ 且 $p_{12}=p, p_{11}=q$. 试证明: 如果 $p>q$，则状态是非常返的；如果 $p=q$，则状态是常返零状态. 试求平稳分布。

9. **两个反射壁**. 设链的状态为 $1, 2, \cdots, a$，其转移概率矩阵的第一行和最后一行分别为 $(q, p, 0, \cdots, 0)$ 与 $(0, \cdots, 0, q, p)$. 在所有其它各行中，$p_{k,k+1}=p$, $p_{k,k-1}=q$. 试求平稳分布. 这个链能否为周期的?

10. 将 N 个黑球与 N 个白球放入两个罐中，使得每个罐有 N 个球. 第一个罐中的黑球数为系统的状态. 每次自每个罐中随机选取一个球，然后将选出的两个球相互交换. 试求 p_{jk}. 证明在极限分布中项 u_k 等于从由 N 个黑球与 N 个白球所组成的总体中任意选出的 N 个球中恰有 k 个黑球的概率[1)].

11. 设链的状态为 $E_0, E_1, \cdots$，转移概率为

$$p_{jk}=e^{-\lambda}\sum_{\nu=1}^{j}\binom{j}{\nu}p^{\nu}q^{j-\nu}\frac{\lambda^{k-\nu}}{(k-\nu)!},$$

此处和中的各项当 $\nu>k$ 时应当用零来代替. 证明

$$p_{jk}^{(n)}\to e^{-\lambda/q}\frac{(\lambda/q)^k}{k!}.$$

注. 这个链在统计力学中遇到[47]，可解如下. 系统的状态由空间某个体积中的质点数来确定. 在每个单位长度的时间间隔中每个质点离开该体积的概率为 q，且各个质点是统计独立的. 此外，新的质点可以进入该体积，且有 r 个新质点进入的概率由普阿松表达式 $e^{-\lambda}\lambda^r/r!$ 给出. 于是平稳分布

---

1) 这个问题要追溯到拉普拉斯；见弗雷谢的书 [38].

为参数为 λ/q 的普阿松分布。

12. **爱伦弗斯特模型**. 在例 (2.f) 中,设初始时第一个容器中有 j 个分子,如果在时刻 n 系统处于状态 k, 则令 $\mathbf{X}^{(n)} = 2k - a$(故 $\mathbf{X}^{(n)}$ 是两个容器中分子数之差)。设 $e_n = \mathbf{E}(\mathbf{X}^{(n)})$. 证明 $e_{n+1} = (a-2)e_n/a$, 因而 $e_n = (1-2/a)^n(2j-a)$. (注意: 当 $n\to\infty$ 时 $e_n\to 0$.)

13. 把第十三章例 (1.b) 中的计数器问题作为一个马尔科夫链来讨论。

14. **具有反射壁的平面随机徘徊**. 考虑在平面有界区域中的对称随机徘徊. 边界在如下意义下是反射的: 每当质点在自由随机徘徊中会离开区域时, 它总被迫使回到原来的位置. 试证明, 如果区域中的每个点都能由每一其它点达到, 则平稳分布存在, 且 $u_k = 1/a$, 此处 a 为区域中的位置数。

15. **累次平均**. 设 $\{x_1, x_2, \cdots\}$ 为一数列且 P 为一个遍历链的矩阵. 试证明 $\sum_j p_{ij}^{(n)}x_j \to \sum u_j x_j$. 证明第十三章第 4 节及第十三章问题 5 中的累次平均程序是一个特殊情况。

16. 在排队论中遇到链矩阵

$$\begin{bmatrix} p_0 & p_1 & p_2 & p_3 & \cdots \\ p_0 & p_1 & p_2 & p_3 & \cdots \\ 0 & p_0 & p_1 & p_2 & \cdots \\ 0 & 0 & p_0 & p_1 & \cdots \end{bmatrix}$$

其中 $\{p_k\}$ 是一概率分布. 利用母函数来讨论状态的特征. 如果平稳分布的母函数存在,试求之。

17. **吸收的等待时间**. 对于非常返状态 E_j 设 $\mathbf{Y}_j$ 为系统第一次转变为常返状态的时间. 假定永远停留于非常返状态的概率为零,证明 $d_j = \mathbf{E}(\mathbf{Y}_j)$ 是如下线性方程组的唯一解:

$$d_j - \sum_T p_{j\nu}d_\nu = 1,$$

其中和是对所有使得 E_ν 为非常返状态的那些 ν 来求的. 不过, d_ν 不必是有限的。

18. 设状态数为 $a<\infty$ 且 E_k 能由 E_j 到达,则它可经不多于 a 步而到达。

19. 设链包含 a 个状态且 E_j 是常返的. 试证明存在一个数 $q<1$, 使得当 $n>a$ 时 E_j 的循环时间超过 n 的概率小于 q^n. (提示: 利用问题 18.)

20. 在有限链中 E_j 为非常返状态的充要条件是, 存在 E_k 使得 E_k 能由

E_j到达，而E_j不能由E_k到达。（如问题7所表明的，本问题中的结论对无限链不成立。）

21．有一个对角线上的元素p_{jj}为正的不可约链不可能是周期的。

22．有限不可约链为非周期链的充要条件是存在n，使得对于所有的j与k有$p_{jk}^{(n)} > 0$。

23．在具有a个状态的链中设$(x_1, \cdots, x_a)$是线性方程组$x_j = \sum p_{j\nu} x_\nu$的一个解．证明：(a) 使得$x_r > 0$的状态构成一个闭(不必不可约)集；(b) 如果$E_j$与$E_k$属于同一不可约集，则$x_j = x_k$。

24．(续上)．如果$(x_1, \cdots, x_a)$是$x_j = s\sum p_{j\nu} x_\nu$的一个解(其中$|s| = 1$但$s \neq 1$)，则存在整数$t > 1$使得$s^t = 1$．如果链是不可约的，则这种整数中的最小者是链的周期。

25．**平均遍历定理**[1]．在任意链中，令

$$A_{jk}^{(n)} = \frac{1}{n} \sum_{\nu=1}^{n} p_{jk}^{(\nu)}.$$

如果E_j与E_k属于同一不可约闭集，则$A_{jk}^{(n)}$趋向于一个极限，此极限与j无关且等于u_k(假定后者存在)．如果E_j与E_k属于不同的闭集，则对于所有的n，$A_{jk}^{(n)} = 0$。如果E_k是非常返的，则对于所有的j，$A_{jk}^{(n)} \to 0$。

26．**移动平均**．设$\{\mathbf{Y}_k\}$为相互独立的随机变量序列，每个变量以概率1/2取值± 1。令$\mathbf{X}^{(n)} = (\mathbf{Y}_n + \mathbf{Y}_{n+1})/2$．求转移概率

$$p_{jk}(m, n) = \mathbf{P}\{\mathbf{X}^{(n)} = k | \mathbf{X}^{(m)} = j\},$$

此处$m < n$且$j, k = -1, 0, 1$．证明$\{\mathbf{X}^{(n)}\}$不是马尔可夫过程且(10.3)不成立。

27．在伯努利试验序列中，如果第$n-1$与n次试验的结果为成功，则说观察到状态E_1；否则我们说系统处于状态E_2．试求n步转移概率并讨论非马尔科夫性。

注．这个过程由问题1中的链通过将三个状态合并为一而得到．这种并合程序可应用于任何马尔科夫链，而使马尔科夫性遭到破坏．在哈里斯的论文中研究过这种过程。

28．**马尔科夫链的混合**．已给状态数目相同的两个马尔科夫链，其转移概率矩阵分别为P_1与P_2．由一个初始分布及n步转移概率$\frac{1}{2}P_1^n + \frac{1}{2}P_2^n$定义一个新过程．试讨论新过程的非马尔科夫性及其与第五章的罐子模型的关系。

---

1) 这个定理是本章结果的一个简单推论．不过，它要简单得多，因此能用较简单的方法来证明；见[48]．

*第十六章　有限马尔科夫链的代数论述

在本章中我们考虑具有有限多个状态 $E_1, \cdots, E_a$ 及一个给定的转移概率矩阵 p_{jk} 的马尔科夫链. 我们的主要目的是导出 n 步转移概率 $p_{jk}^{(n)}$ 的明显公式. 除了第 3 节中的一般概念和记号外，我们并不需要上一章中的结果.

我们将利用母函数的方法并将由第十一章第 4 节中的部分分式的展开得出所要求的结果. 我们的结果也能直接由矩阵的典型分解理论[1](反之，它们也能由我们的结果导出)得到. 此外，对于有限链，第十五章中所证明的遍历性质可由本章的结果推出. 不过，为简单起见，我们将对普遍性稍加限制，而不考虑那些使一般理论复杂化而在实际例题中不会出现的特殊情况.

在第 1 节中概述了一般方法，例题在 2、3 两节中给出. 在第 4 节中着重讨论了非常返状态与吸收概率. 在第 5 节中理论被用来计算状态 E_j 的循环时间的方差.

1. 一 般 理 论

对于每个固定的 j, k 我们定义一个母函数

$$P_{jk}(s) = \sum_{n=1}^{\infty} p_{jk}^{(n)} s^{n-1}. \tag{1.1}$$

以 $sp_{\nu j}$ 乘上式并对 j 求和，我们得

$$s \sum_{j=1}^{a} p_{\nu j} P_{jk}(s) = P_{\nu k}(s) - p_{\nu k}. \tag{1.2}$$

---

*本章讨论专门的论题，可以略去.

1) 见弗雷谢的论文 [38].

对于每个固定的 k，这里我们有由关于 a 个未知数 $P_{1k}(s)$, $\cdots$, $P_{ak}(s)$ 的 a 个非齐次线性方程所构成的方程组．理论上，这个方程组能用行列式或逐次消元法来解．我们仅利用如下事实：方程组的行列式 $D(s)$是次数不超过 a 的多项式,而 $P_{\nu k}(s)$则为以 $D(s)$ 为公分母的 s 的有理函数．我们仅考虑方程 $D(s)=0$ 无重根的情况;这对一般性是一个轻微的限制,但我们的理论包括了具有实用价值的大多数情况．

因为 $P_{\nu k}(s)$ 是有理函数，由第十一章第 4 节的部分分式的展开可知,存在系数 $\rho_{\nu k}^{(1)}$, $\cdots$, $\rho_{\nu k}^{(a)}$, 使得

$$p_{\nu k}^{(n)}=\frac{\rho_{\nu k}^{(1)}}{s_1^n}+\frac{\rho_{\nu k}^{(2)}}{s_2^n}+\cdots+\frac{\rho_{\nu k}^{(a)}}{s_a^n}, \tag{1.3}$$

此处 s_1, s_2, $\cdots$ 是 $D(s)=0$ 的根．如果 $D(s)$ 的次数小于 a, 则(1.3)所包含的项数小于 a．对于 ν 与 k 的某些特殊值，分子与分母有一个或多个根 s_r 相同因而相互消去也是可能的．在这种情况我们令对应的 $\rho_{\nu k}^{(r)}$ 为零．

根 s_r 与系数 $\rho_{\nu k}^{(r)}$ 可用第十一章中的方法来计算，但利用马尔科夫链的某些特殊性质来计算更好一些．以 $p_{j\nu}$ 乘方程(1.3)并对 $\nu=1, 2, \cdots$ 求和,得

$$p_{jk}^{(n+1)}=\sum_{\nu=1}^{a} p_{j\nu}\left\{\frac{\rho_{\nu k}^{(1)}}{s_1^n}+\cdots+\frac{\rho_{\nu k}^{(a)}}{s_a^n}\right\}. \tag{1.4}$$

如果用(1.3)表示上式左边,则我们得到一个恒等式，仅当两边 s_1^{-n}, $\cdots$, s_a^{-n} 的系数相等时，它才能对所有的 n 成立．这意味着,对于固定的 r，我们必须有

$$\rho_{jk}^{(r)}=s_r\sum_{\nu=1}^{a} p_{j\nu}\rho_{\nu k}^{(r)},\ r=1, \cdots, a. \tag{1.5a}$$

类似,以 p_{km} 乘(1.3)两边并对所有的 k 求和,我们得

$$\rho_{\nu k}^{(r)}=s_r\sum_{k=1}^{a} \rho_{\nu k}^{(r)}p_{km}. \tag{1.5b}$$

关系式(1.5a)表明,对于固定的 k 与 r，a 个量 $\rho_{1k}^{(r)}$, $\cdots$, $\rho_{ak}^{(r)}$ 是如下 a 元线性方程组的解

$$x_j^{(r)} = s_r \sum_{\nu=1}^{a} p_{j\nu} x_\nu^{(r)} \quad (j = 1, \cdots, a). \tag{1.6a}$$

类似，由关系式 (1.5b) 可知，对于固定的 ν 与 r，$\rho_{\nu 1}^{(r)}, \cdots, \rho_{\nu a}^{(r)}$ 满足 a 个线性方程

$$y_m^{(r)} = s_r \sum_{k=1}^{a} y_k^{(r)} p_{km} \quad (m = 1, \cdots, a). \tag{1.6b}$$

为了更好的理解，我们用任意的 s 来代替 s_r 并研究如下两个更为一般的方程组

$$x_j = s \sum_{\nu=1}^{a} p_{j\nu} x_\nu \quad (j = 1, \cdots, a) \tag{1.7a}$$

及

$$y_m = s \sum_{k=1}^{a} y_k p_{km} \quad (m = 1, \cdots, a). \tag{1.7b}$$

仅当其行列式为零时，a 元齐次方程组才能有非平凡[1]解. 现在方程组 (1.7a) 与 (1.7b) 的矩阵除行列互换外是相同的. 所以它们的行列式相等. 此外，(1.7a) 的行列式显然等于方程组 (1.2) 的行列式，这意味着当 $s = s_1, s_2, \cdots, s_a$ 时，两个方程组 (1.7a) 与 (1.7b) 的行列式为零.

现在我们可以不考虑母函数 $P_{jk}(s)$ 而将根 s_r 定义为使得方程组 (1.7a) 与 (1.7b) 有非平凡解的那些数(实的或复的). s_r 为单根的假设意味着，对于每个固定的 r，除一个数值因子外解 $(x_1^{(r)}, \cdots, x_a^{(r)})$ 与 $(y_1^{(r)}, \cdots, y_a^{(r)})$ 是唯一确定的. 不过，我们的出发点是，对于固定的 k 与 r，$(\rho_{1k}^{(r)}, \cdots, \rho_{ak}^{(r)})$ 是 (1.7a) 的一个解，而对于固定的 ν 与 r，$(\rho_{\nu 1}^{(r)}, \cdots, \rho_{\nu a}^{(r)})$ 是 (1.7b) 的一个解. 因为这些解确定到一个数值因子，故我们必须有

$$\rho_{jk}^{(r)} = c_r x_j^{(r)} y_k^{(r)}. \tag{1.8}$$

剩下只需计算常数 $c_1, \cdots, c_a$.

由 (1.8) 与 (1.3) 我们有

---

1) 照例，我们称恒等于零的解为平凡解，对于它我们不予考虑.

$$\rho_{jk}^{(n)} = \sum_{r=1}^{a} c_r x_j^{(r)} y_k^{(r)} s_r^{-n}. \tag{1.9}$$

所以

$$P_{jk}(s) = c_1 \frac{x_j^{(1)} y_k^{(1)}}{s_1 - s} + \cdots + c_a \frac{x_j^{(a)} y_k^{(a)}}{s_a - s}. \tag{1.9a}$$

利用 (1.6a) 由 (1.1) 可得

$$\sum_{k=1}^{a} P_{jk}(s) x_k^{(r)} = x_j^{(r)} \sum_{n=1}^{\infty} s^{n-1} s_r^{-n} = \frac{x_j^{(r)}}{s_r - s}. \tag{1.10}$$

另一方面，如果我们用 (1.9a) 来求左边的值，则我们导致分母为 $s_\nu - s$ 的 a 个分式的一个和. 由此推得，当 $\nu \neq r$ 时分子必须为零且

$$1 = c_r \sum_{\nu=1}^{a} x_\nu^{(r)} y_\nu^{(r)}, \tag{1.11}$$

于是我们已求得 c_r. 我们已知，解 $x_\nu^{(r)}$ 与 $y_\nu^{(r)}$ 仅确定到一个数值因子. 不过，如果我们用 $Ax_j^{(r)}$ 代替 $x_j^{(r)}$，用 $By_k^{(r)}$ 代替 $y_k^{(r)}$，则 c_r 将变成 c_r/AB 而 (1.8) 中的量则保持不变.

综上所述，我们得到如下的计算 $p_{jk}^{(n)}$ 的程序：写出两个线性方程组 (1.7a) 与 (1.7b). 它们有共同的行列式且仅对 s 的使这个行列式为零的值有非平凡解. 我们假定，根 $s_1, s_2, \cdots$（最多有 a 个）都是单重的；于是对于每个 r，解 $(x_1^{(r)}, \cdots, x_a^{(r)})$ 与 $(y_1^{(r)}, \cdots, y_a^{(r)})$ 确定到一个任意的乘性常数. 求出这些解并由 (1.11) 求出常数 c_r. 于是 $p_{jk}^{(n)}$ 由 (1.9) 给出.

对于每个固定的 r，$\rho_{jk}^{(r)}$ 构成一个矩阵，它可以用如下的方法来构造. 作出一个以 $x_j^{(r)}$ 为头一行、以 $y_k^{(r)}$ 为头一列的乘法表. 用 c_r 乘这个表中的所有 a^2 个元素即得矩阵 $\rho_{jk}^{(r)}$. 为了构造矩阵 $(p_{jk}^{(n)})$，我们必须用 s_r^n 除 $\rho_{jk}^{(r)}$ 的所有元素，并将这样得到的矩阵对 $r = 1, 2, \cdots, a$ 求和. 注意，即使根的个数小于 a，根 s_r 也可能是单重的.

重根的情况需要作某些改变但可用类似的方法来处理. 最重要的情况将在第 4 节中讨论.

在代数中，倒数 $\lambda_r = 1/s_r$ 称为矩阵 P 的特征值(也称固有值或特征根)。零是一个可能的特征值，但没有根 s_r 与之对应。这说明了为什么即使总有 a 个特征值而根 s_r 的数目也可以少于 a。对于母函数的方法来说，使用 s_r 比使用它们的倒数更方便一些。此外，它与积分方程论的一般习惯相对应，因而在概率论中更自然一些。

值 $s = 1$ 总在 s_r 中出现，且(1.7a)的解$(1, 1, \cdots, 1)$与之对应。对于所有的 r，我们有 $|s_r| \geqslant 1$。事实上[1]，如果 $|s_r| < 1$，则根 $|s_r|$ 在(1.3)中会引起发散。如果 $s = 1$ 是使得 $|s_r| = 1$ 的仅有根，则 $p_{jk}^{(n)} \to c_1 x_j^{(1)} y_k^{(1)}$。容易证明，如果存在满足条件 $|s_r| = 1$ 的其它根，则它们必定是1的 t 次根，此处 t 为整数；在这种情况，链具有周期 t。详细的讨论读者可参看弗雷谢的论文[38]。

通常要求出所有的根是麻烦的或者是不可能的。不过，$p_{jk}^{(n)}$ 渐近性态的一阶近似显然由满足条件 $|s_r| = 1$ 的 s_r 所确定，二阶近似则由具有下一个最小绝对值的根 s_r 所确定。

最后的公式(1.9)能用简洁的矩阵记号来表示。设 $X^{(r)}$ 是其元素为 $x_j^{(r)}$ 的列向量(或 $a \times 1$ 矩阵)，又设 $Y^{(r)}$ 是其元素为 $y_k^{(r)}$ 的行向量(或 $1 \times a$ 矩阵)。则 $X^{(r)}Y^{(r)}$ 是其元素为 $x_j^{(r)}y_k^{(r)}$ 的 $a \times a$ 矩阵，且(1.9)可写为如下形式

$$P^n = \sum_{r=1}^{a} c_r s_r^{-n} X^{(r)} Y^{(r)} \quad \text{其中 } c_r^{-1} = Y^{(r)} X^{(r)}. \tag{1.12}$$

向量 $X^{(r)}$ 与 $Y^{(r)}$ 称为特征向量，c_r^{-1} 为它们的内积。

2. 例　题

(a) 首先考虑仅有两个状态的链。转移概率矩阵具有如下的简单形式

$$P = \begin{pmatrix} 1-p & p \\ \alpha & 1-\alpha \end{pmatrix}$$

其中 $0 < p < 1$，$0 < \alpha < 1$。方程(1.7a)化归为 $s(1-p)x_1 + spx_2 = x_1$ 及 $s\alpha x_1 + s(1-\alpha)x_2 = x_2$。使两个比值 x_1/x_2 相等，可

1) 直接的证明如下。设 M 为序列 $|x_1^{(r)}|, \cdots, |x_a^{(r)}|$ (r 固定)中的最大项。则由(1.6a)有 $M \leqslant |s_r| \sum p_{j\nu} M = |s_r| M$ 或 $|s_r| \geqslant 1$。

以看出仅当 $s=1$ 或 $s=1/(1-\alpha-p)$ 时存在一个解. 对应于 $s_1=1$ 的解为(1,1); 对应于 $s_2=1/(1-\alpha-p)$ 的解为 $(p, -\alpha)$. 其次考虑方程组(1.7b), 现在它化归为 $s(1-p)y_1+s\alpha y_2=y_1$ 及 $spy_1+s(1-\alpha)y_2=y_2$. 我们知道, 仅当 $s=s_1$ 或 $s=s_2$ 时它才能解出. 对应的解为 (α, p) 与 $(1, -1)$. 由(1.11)我们得 $c_1=c_2=1/(\alpha+p)$. 现在方程(1.9)与(1.11)使我们能够写出量 $p_{jk}^{(n)}$ 的明显公式. 最后的结果能写成如下的矩阵形式

$$P^n=\frac{1}{\alpha+p}\begin{pmatrix}\alpha & p\\ \alpha & p\end{pmatrix}+\frac{(1-\alpha-p)^n}{\alpha+p}\begin{pmatrix}p & -p\\ -\alpha & \alpha\end{pmatrix}$$

(此处已将四个元素的公因子提出作为矩阵的因子). 因为 $|1-\alpha-p|<1$, 故当 $n\to\infty$ 时第二个矩阵趋向于零, 而第一个矩阵则表示 P^n 的极限形式.

(b) 设

$$P=\begin{bmatrix}0 & 0 & 0 & 1\\ 0 & 0 & 0 & 1\\ \frac{1}{2} & \frac{1}{2} & 0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix} \tag{2.1}$$

[这是第十五章问题 2(b) 的矩阵]. 方程组(1.7a)化归为

$$x_1=sx_4,\qquad x_2=sx_4,$$
$$x_3=\frac{s(x_1+x_2)}{2},\qquad x_4=sx_3. \tag{2.2}$$

因为乘性常数任意, 故我们可令 $x_4=1$. 于是 $x_1=s$, $x_2=s$, $x_3=s^2$, $x_4=s^3$, 因此我们必须有 $s^3=1$. 现在如果我们令

$$\theta=e^{2\pi i/3}=\cos\frac{2\pi}{3}+i\sin\frac{2\pi}{3}, \tag{2.3}$$

则 $s^3=1$ 的三个根为 $s_1=1$, $s_2=\theta$, $s_3=\theta^2$. (注意: 即使有四个状态, 我们也仅有三个根.) 对应于这三个根的解为 $(1, 1, 1, 1)$, $(\theta, \theta, \theta^2, 1)$, $(\theta^2, \theta^2, \theta, 1)$.

由方程组(1.7b)我们得 $y_1=sy_3/2$, $y_2=sy_3/2$, $y_3=sy_4$,

$y_4 = s(y_1 + y_2)$. 对应于 $s_1 = 1$, $s_2 = \theta$ 与 $s_3 = \theta^2$ 的三组解为 $(1, 1, 2, 2)$, $(\theta, \theta, 2, 2\theta^2)$, $(\theta^2, \theta^2, 2, 2\theta)$. 所以由 (1.11) 有 $c_1 = 1/6$, $c_2 = 1/(6\theta^2) = \theta/6$, $c_3 = 1/(6\theta) = \theta^2/6$. 现在我们能够表示所有的 $p_{jk}^{(n)}$. 例如,

$$p_{11}^{(n)} = p_{22}^{(n)} = \frac{1 + \theta^n + \theta^{2n}}{6},$$

$$p_{13}^{(n)} = \frac{1 + \theta^{2n+2} + \theta^{n+1}}{3}, \tag{2.4}$$

$$p_{14}^{(n)} = \frac{1 + \theta^{2n+1} + \theta^{n+2}}{3},$$

等等. 这个链显然是周期的,其周期为 3.

(c) 设 $p + q = 1$,

$$P = \begin{bmatrix} 0 & p & 0 & q \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ p & 0 & q & 0 \end{bmatrix}. \tag{2.5}$$

[这个矩阵描述了循环随机徘徊;见第十五章例(2.d).] 方程(1.7a)化归为 $x_1 = s(px_2 + qx_4)$, $x_2 = s(qx_1 + px_3)$, $x_3 = s(qx_2 + px_4)$, $x_4 = s(px_1 + qx_3)$. 假定 $p \neq q$. 由第一个与第三个方程我们求得 $x_1 + x_3 = s(x_2 + x_4)$, 由其余两个方程有 $x_2 + x_4 = s(x_1 + x_3)$. 所以我们有 $s^2 = 1$ 或 $x_1 + x_3 = x_2 + x_4 = 0$. 在第一种情况有两个根 $s_1 = 1$, $s_2 = -1$. 另一方面,以 $x_3 = -x_1$, $x_4 = -x_2$ 代入前两个方程,我们有 $s^2(p - q)^2 = -1$, 由此可得另两个根 s_3 与 s_4. 于是

$$s_1 = 1,\ s_2 = -1,\ s_3 = \frac{i}{q - p},\ s_4 = -\frac{i}{q - p}, \tag{2.6}$$

(此处 $i^2 = -1$). 对应的解 $x_j^{(r)}$ 包含一个任意的因子,故我们可令 $x_4^{(r)} = 1$. 于是容易求得四组解为 $(1, 1, 1, 1)$, $(-1, 1, -1, 1)$, $(i, -1, -i, 1)$, $(-i, -1, i, 1)$. 在我们的情况方程组(1.7b)化归为 $y_1 = s(qy_2 + py_4)$, $y_2 = s(py_1 + qy_3)$, $y_3 = s(py_2 +$

qy_4), $y_4 = s(qy_1 + py_3)$. 对应于(2.6)中的四个根的解为(1, 1, 1, 1), (−1, 1, −1, 1), (−i, −1, i, 1), (i, −1, −i, 1). 对于常数 c_r 我们由 (1.11) 求得 $c_1 = c_2 = c_3 = c_4 = 1/4$. 利用 (1.3) 与 (1.8), 现在对于每个序列 $p_{jk}^{(n)}(n = 1, 2, 3, \cdots)$ 我们能写出一个明显公式. 在现在的情况,解 $x_j^{(r)}$ 与 $y_j^{(r)}$ 具有简单的形式 $(\alpha, \alpha^2, \alpha^3, \alpha^4)$, 此处 α 是四个数 1, −1, i 或 −i 之一. 这使我们能够用简单的公式来表示 $p_{jk}^{(n)}$:

$$p_{jk}^{(n)} = \frac{1}{4}\{1 + (q - p)^n (i)^{j-k-n}\}\{1 + (-1)^{k+j-n}\}. \tag{2.7}$$

这个公式对于 $p = q = 1/2$ 也成立.

可以看出, 包含 $(q - p)^n$ 的项趋向于零,其它的项则具有周期 2.

(d) 一般的循环随机徘徊 [第十五章例 (2.d)]. 在上个例子中我们能将 $x_j^{(r)}$ 与 $y_k^{(r)}$ 表示为 1 的四个四次根的幂. 这启发我们对第十五章例 (2.d) 中的一般矩阵试用类似的程序. 将状态从 0 到 $a - 1$ 来编号是便利的. 为简洁起见,我们令

$$\theta = e^{2i\pi/a}. \tag{2.8}$$

这是 1 的一个 a 次根, 且所有的 a 次根可用序列 $1, \theta, \theta^2, \cdots, \theta^{a-1}$ 来表示. 容易看出,方程组 (1.7a) 与 (1.7b) 有如下 a 组解

$$x_j^{(r)} = \theta^{rj}, \quad y_k^{(r)} = \theta^{-rk}, \tag{2.9}$$

其中 $r = 0, 1, 2, \cdots, a - 1$; 它们对应于

$$s_r = \left\{\sum_{\nu=1}^{a-1} q_\nu \theta^{\nu r}\right\}^{-1}. \tag{2.10}$$

由方程(1.11)与(2.9)可知, 对于一切 r 有 $c_r = 1/a$, 于是最后有

$$p_{jk}^{(n)} = \frac{1}{a}\sum_{r=0}^{a-1} \theta^{r(j-k)}\left(\sum_{\nu=0}^{a-1} q_\nu \theta^{\nu r}\right)^n. \tag{2.11}$$

对 $n = 1$ 来验证这个公式是有趣的. q_ν 的因子为

$$\sum_{r=0}^{a-1} \theta^{r(j-k+\nu)}. \tag{2.12}$$

当 $j-k+\nu=0$ 或 a 时,和中的每一项等于 1;对于所有其它情况,此和为零. 于是,如果 $k \geqslant j$, 则 $p_{jk}^{(1)}$ 化归为 q_{k-1}; 如果 $k<j$, 则 $p_{jk}^{(1)}$ 化归为 q_{a+k-j}, 而这就是所给的矩阵 (p_{jk}).

(e) **占位问题**. 第十五章例 (2.g) 表明,古典占位问题能用马尔科夫链的方法来处理. 如果有 j 个盒被占据、$a-j$ 个盒空着,则系统处于状态 j. 如果这是初始情况,而另有 n 个球被随机地放入盒中, 则 $p_{jk}^{(n)}$ 为将有 k 个盒被占据、$a-k$ 个盒空着的概率(因此如果 $k<j$, 则 $p_{jk}^{(n)}=0$). 对于 $j=0$ 这个概率由第二章中的公式(11.7)得到. 我们现在导出 $p_{jk}^{(n)}$ 的公式, 因而推广第二章的结果.

因为 $p_{jj}=j/a$, $p_{j,j+1}=(a-j)/a$, 故容易看出, 方程组 (1.7a) 化归为

$$(a-sj)x_j=s(a-j)x_{j+1}, \quad j=0, \cdots, a. \tag{2.13}$$

对于 $s=1$ 我们得到解 $x_j=1$. 显然,如果 $s \neq 1$, 则 $x_a=0$, 故 $s=1$ 是使得所有的 x_j 异于零的 s 的唯一值. 如果 s 是使得 (2.13) 有解的任一其他值, 则必须存在某个足标 $r<a$, 使得 $x_{r+1}=0$ 而 $x_r \neq 0$; 于是由 (2.13) 可得 $sr=a$. 因而使得 (2.13) 有解的根 s_r 为 $s_r=a/r(r=1, 2, \cdots, a)$. (2.13) 的对应解可依次令 $x_0^{(r)}=1$ 及 $j=0, 1, \cdots$ 而得. 我们求得

$$x_j^{(r)}=\binom{r}{j} \div \binom{a}{j}, \tag{2.14}$$

故当 $j>r$ 时 $x_j^{(r)}=0$.

对于 $s=s_r$ 方程组 (1.7b) 化归为

$$(r-j)y_j^{(r)}=(a-j+1)y_{j-1}^{(r)}, \tag{2.15}$$

其解为

$$y_j^{(r)}=\binom{a-r}{j-r}(-1)^{j-r}, \tag{2.16}$$

这里,当然当 $j<r$ 时 $y_j^{(r)}=0$. 因为当 $j>r$ 时 $x_j^{(r)}=0$ 而当 $j<r$ 时 $y_j^{(r)}=0$, 故由方程(1.11)我们容易求出 $c_r=(x_r^{(r)}y_r^{(r)})^{-1}$

$=\binom{a}{r}$，因此

$$p_{jk}^{(n)}=\sum_{r=j}^{k}\left(\frac{r}{a}\right)^n\binom{a}{r}\binom{r}{j}\binom{a-r}{k-r}(-1)^{k-r}\div\binom{a}{j}. \quad (2.17)$$

用阶乘来表示二项式系数，这个公式可简化为

$$p_{jk}^{(n)}=\binom{a-j}{a-k}\sum_{\nu=0}^{k-j}\left(\frac{\nu+j}{a}\right)^n(-1)^{k-j-\nu}\binom{k-j}{\nu}, \quad (2.18)$$

如果 $k<j$，则 $p_{jk}^{(n)}=0$.

（进一步的例题可在以下两节中找到.）

3. 具有反射壁的随机徘徊

现在我们通过详细讨论具有状态 $1, 2, \cdots, a$ 及两个反射壁的随机徘徊来说明马尔科夫链的应用[1]. 矩阵 P 的第 $2, 3, \cdots, a-1$ 行由 $p_{k,k+1}=p$ 及 $p_{k,k-1}=q$ 所确定；第一行和最后一行分别为 $(q, p, 0, \cdots, 0)$，$(0, \cdots, 0, q, p)$. 当 $\delta=1$ 时第十五章例 (2.c) 的矩阵简化为这个矩阵. 用随机徘徊的术语来说，$p_{jk}^{(n)}$ 是由 $x=j$ 出发的质点在时刻 n 位于 $x=k$ 的概率.

方程 (1.7a) 取如下形式

$$\begin{aligned}x_1&=s(qx_1+px_2)\\x_j&=s(qx_{j-1}+px_{j+1} \quad (j=2, 3, \cdots, a-1)\\x_a&=s(qx_{a-1}+px_a).\end{aligned} \quad (3.1)$$

这个方程组对应于 $s=1$ 的解为 $x_j\equiv1$. 为求所有其它的解，我们应用特解方法(在第十四章第 4 节中我们已将这个方法应用于类似的方程). 如果 λ 是二次方程 $\lambda=qs+\lambda^2ps$ 的一个根，则 (3.1) 的中间的方程被 $x_j=\lambda^j$ 所满足. 这个方程的两个根为

---

1) 以下的部分讨论是第十四章中的理论的重复. 我们的二次方程在第十四章中作为 (4.7) 而出现；量 $\lambda_1(s)$ 与 $\lambda_2(s)$ 曾由 (4.8) 给出，一般解 (3.3) 在第十四章中则作为 (4.9) 而出现. 这两种方法是有关联的，但在很多地方计算的细节根本不同.

$$\lambda_1(s) = \frac{1 + (1 - 4pqs^2)^{\frac{1}{2}}}{2ps}, \quad \lambda_2(s) = \frac{1 - (1 - 4pqs^2)^{\frac{1}{2}}}{2ps}, \tag{3.2}$$

所以(3.1)中的中间的方程的最一般解为

$$x_j = A(s)\lambda_1^j(s) + B(s)\lambda_2^j(s), \tag{3.3}$$

此处 $A(s)$ 与 $B(s)$ 是任意的．(3.3)满足(3.1)中的第一个与最后一个方程的充要条件是 $x_0 = x_1$ 且 $x_a = x_{a+1}$. 这需要 $A(s)$ 与 $B(s)$ 满足如下条件

$$\begin{aligned} &A(s)\{1 - \lambda_1(s)\} + B(s)\{1 - \lambda_2(s)\} = 0, \\ &A(s)\lambda_1^a(s)\{1 - \lambda_1(s)\} + B(s)\lambda_2^a(s)\{1 - \lambda_2(s)\} = 0. \end{aligned} \tag{3.4}$$

不过，仅当

$$\lambda_1^a(s) = \lambda_2^a(s) \tag{3.5}$$

时，这两个方程才是相容的，于是我们必须确定 s 的值使(3.5)成立．

由定义(3.2)我们有 $\lambda_1(s)\lambda_2(s) = q/p$，由(3.5)可知，$\lambda_1(s)(p/q)^{\frac{1}{2}}$ 与 $\lambda_2(s)(p/q)^{\frac{1}{2}}$ 必须是1的 $2a$ 次根．这些根能写成如下形式

$$e^{i\pi r/a} = \cos\frac{\pi r}{a} + i\sin\frac{\pi r}{a}, \tag{3.6}$$

这里 $i^2 = -1$，$r = 0, 1, 2, \cdots, 2a-1$. 于是(3.5)的所有解在

$$\lambda_1(s) = \left(\frac{q}{p}\right)^{\frac{1}{2}} e^{i\pi r/a}, \quad \lambda_2(s) = \left(\frac{q}{p}\right)^{\frac{1}{2}} e^{-i\pi r/a}$$

的根之中．

对于每个值 r 我们能求得一个根 s_r，即

$$s_r = \{2(pq)^{\frac{1}{2}} \cos \pi r/a\}^{-1}. \tag{3.7}$$

值 $r = a$ 必须除外，因为对于它 $\lambda_1(s) = \lambda_2(s)$，$A(s) = -B(s)$，故它导致平凡解 $x_j \equiv 0$. 对应于 $r = 0$ 的解为 $x_j \equiv 1$，这个解我们已予考虑．对于 $r = 1, 2, \cdots, a-1$，有 $a-1$ 个不同的解与之对应；如果我们令 $r = a+1, a+2, \cdots, 2a-1$，则我们

得到相同的解，其中 $\lambda_1(s)$ 与 $\lambda_2(s)$ 互换。这样我就已求得 (3.1) 的 a 组不同的解，且我们知道，不可能有更多的解.

对于 $s = s_r(r = 1, 2, \cdots, a-1)$，由 (3.4) 可得 $2A(s) = 1-\lambda_2(s)$ 及 $2B(s) = -\{1-\lambda_1(s)\}$.（记住：乘性常数保持任意.）代入 (3.3)，我们求得 $a-1$ 组解

$$x_j^{(r)} = \left(\frac{q}{p}\right)^{\frac{1}{2}j} \sin\frac{\pi rj}{a} - \left(\frac{q}{p}\right)^{\frac{1}{2}(j+1)} \sin\frac{\pi r(j-1)}{a} \tag{3.8}$$

$(r = 1, 2, \cdots, a-1)$. 将前面已求得的解

$$x_j^{(0)} = 1 \tag{3.9}$$

添加到这组解中. 容易验证 (3.8) 与 (3.9) 表示已给方程组 (3.1) 的解.

现在我们必须求第二组线性方程的解. 在现在的情况，(1.7b) 取如下形式

$$\begin{aligned} y_1 &= sq(y_1 + y_2), \\ y_k &= s(py_{k-1} + qy_{k+1}), \quad (k = 2, \cdots, a-1) \\ y_a &= sp(y_{a-1} + y_a). \end{aligned} \tag{3.10}$$

将 p 与 q 互换后中间的方程与 (3.1) 是一样的，故它的一般解可由交换 (3.3) 中的 p 与 q 而得. 如果 $s = s_r$，则第一个与最后一个方程能被满足，用简单的计算即可证明，对于 $r = 1, 2, \cdots, a-1$，(3.10) 的解为

$$y_k^{(r)} = \left(\frac{p}{q}\right)^{\frac{1}{2}k} \sin\frac{\pi rk}{a} - \left(\frac{p}{q}\right)^{\frac{1}{2}(k-1)} \sin\frac{\pi r(k-1)}{a}. \tag{3.11}$$

对于 $s = 1$ 可类似地求出

$$y_k^{(0)} = \left(\frac{p}{q}\right)^k. \tag{3.12}$$

下一步在于计算 (1.11) 中的系数 c_r. 如用倍角的余弦来表示 $\sin^2 \pi rj/a$ 并利用复指数，则求和可以简化. 于是我们只须求有限几何级数之和，容易求得

$$c_r = \frac{2p}{a}\left\{1 - 2(pq)^{\frac{1}{2}} \cos\frac{\pi r}{a}\right\}^{-1} \quad (r = 1, 2, \cdots, a-1). \tag{3.13}$$

对于 $r=0$，如果 $p\neq q$，则我们得

$$c_0=\frac{q}{p}\frac{(p/q)-1}{(p/q)^a-1}. \tag{3.14}$$

如果 $p=q=1/2$，则(3.13)仍保持有效，但(3.14)要用 $c_0=1/a$ 来代替.

这些公式导致最后的结果

$$p_{jk}^{(n)}=\frac{(p/q)-1}{(p/q)^a-1}\left(\frac{p}{q}\right)^{k-1}+\frac{2^{n+1}p^{1+\frac{1}{2}(n-j+k)}q^{\frac{1}{2}(n+j-k)}}{a}\sum_{r=1}^{a-1}S_r, \tag{3.15}$$

其中 S_r 代表

$$\frac{\cos^n\frac{\pi r}{a}\left\{\sin\frac{\pi rj}{a}-\left(\frac{q}{p}\right)^{\frac{1}{2}}\sin\frac{\pi r(j-1)}{a}\right\}\left\{\sin\frac{\pi rk}{a}-\left(\frac{q}{p}\right)^{\frac{1}{2}}\sin\frac{\pi r(k-1)}{a}\right\}}{1-2(pq)^{\frac{1}{2}}\cos\frac{\pi r}{a}}.$$

当 $n\to\infty$ 时，(3.15)中的第二项趋向于零，故 $p_{jk}^{(n)}$ 趋向于不依赖于 j 的平稳分布.（这个极限分布可用其它的方法导出，参见第十五章问题 9.）令 $a\to\infty$ 并取极限，我们就得到具有单一反射壁的随机徘徊的公式；在取极限时，和式被一个积分来代替[1].

4. 非常返状态；吸收概率

第 1 节的定理是在根 s_1，s_2，$\cdots$ 互不相同的假设下导出的. 对于存在重根的情况并不需作本质的修改，但我们仅讨论特别重要的一个特殊情况. 每当链包含两个或多个闭子链时，$s_1=1$ 是重根，这种情况在与吸收概率有关的问题中经常出现. 容易修改第 1 节中的方法使它适应于这种情况. 为简单明了起见，我们将

---

1) 对于在具有一个反射壁与一个吸收壁的情况的类似公式见 M. 卡斯 [49]. 在那里对反射壁的概念作了修改，使得质点可以到达 0. 每当这种情况出现时，质点在下一步就回到 1. 在这种情况，明显的公式比较复杂. 卡斯也求出了爱伦弗斯特模型[第十五章例 (2.f)] 中关于 $p_{jk}^{(n)}$ 的公式.

通过例题来说明这个程序，这些例题将揭示一般情况的主要特征。

例．(a)**考虑转移概率矩阵**

$$P=\begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 & 0 \\ \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{5} & \frac{4}{5} & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix}. \tag{4.1}$$

显然，E_1 与 E_2 构成一个闭集(即不可能由它们转移到其它四个状态中的任何一个；试与第十五章第 4 节相比较)．类似地，E_3 与 E_4 构成另一闭集．最后，E_5 与 E_6 为非常返状态． 在有限步以后系统进入两个闭集之一并停留在其中．

矩阵 P 具有分块矩阵的形式：

$$P^n=\begin{bmatrix} A & 0 & 0 \\ 0 & B & 0 \\ U & V & T \end{bmatrix}, \tag{4.2}$$

此处每个字母代表一个二阶矩阵，每个零代表由四个零组成的矩阵．例如，A 具有行 $(\frac{1}{3}, \frac{2}{3})$ 与 $(\frac{2}{3}, \frac{1}{3})$；这是对应于由两个状态 E_1 与 E_2 所构成的链的转移概率矩阵．这个矩阵可单独地加以研究，而幂 A^n 则可由在例 (2.a) 中令 $p=\alpha=2/3$ 而得到．在计算幂 P^2, P^3, $\cdots$ 时，我们将看到，前两行决不受其余四行的影响．更精确地说，P^n 具有形式

$$P=\begin{bmatrix} A^n & 0 & 0 \\ 0 & B^n & 0 \\ U_n & V_n & T^n \end{bmatrix}, \tag{4.3}$$

这里 A^n, B^n, T^n 分别为 A, B 与 T 的 n 次幂，且能用第 1 节的方法来计算[1)]（参见例 (2.a)，在那里所有的计算都已完成）. 代替六个六元方程，我们所面临的是各包含两个方程的二元方程组.

应当指出，(4.3) 中的矩阵 U_n 与 V_n 不是 U 与 V 的幂，因而不能用像求 A^n, B^n 与 T^n 那样简单的方法得到. 在计算 P^2, P^3, $\cdots$ 时，第三、四两列决不影响其余四列. 换句话说，如果在 P^n 中将对应于 E_3 与 E_4 的行与列删去，则我们得到矩阵

$$\begin{pmatrix} A^n & 0 \\ U_n & T^n \end{pmatrix}. \tag{4.4}$$

它是 P 中的对应子矩阵

$$\begin{pmatrix} A & 0 \\ U & T \end{pmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix} \tag{4.5}$$

的 n 次幂. 所以矩阵 (4.4) 能用第 1 节中的方法来计算，在现在的情况这种方法大为简化. 矩阵 V_n 可用类似的方法得到.

通常 U_n 与 V_n 的明显形式仅由于它们与吸收概率有关才具有重要性. 设系统从，比方说，E_5 出发，试问：它最终将进入由 E_1 与 E_2 所构成的闭集（因而不进入其它闭集）的概率 λ 是什么？这个事件恰好在第 n 步出现的概率 λ_n 是什么？显然，$p_{51}^{(n)} + p_{52}^{(n)}$ 是所考虑的事件在第 n 步或在此以前出现的概率，即

$$p_{51}^{(n)} + p_{52}^{(n)} = \lambda_1 + \lambda_2 + \cdots + \lambda_n.$$

令 $n \to \infty$，即得 λ. 计算 λ_n 的一个较好的方法如下. 第 $n-1$ 步必须使系统处于异于 E_1 与 E_2 的一个状态，即处于 E_5 或 E_6（因为

---

1) 在 T 中行和不为 1，故 T 不是随机矩阵. 不过，除 $s=1$ 不再是一个根（故 $T^n \to 0$）外，第 1 节中的方法无需改变就可应用.

由 E_3 或 E_4 不可能转移到 E_1 与 E_2). 然后第 n 步使系统处于 E_1 或 E_2. 所以

$$\lambda_n = p_{55}^{(n-1)}(p_{51} + p_{52}) + p_{56}^{(n-1)}(p_{61} + p_{62})$$
$$= \frac{1}{4}\ p_{55}^{(n-1)} + \frac{1}{3}\ p_{56}^{(n-1)}. \tag{4.6}$$

λ_n 完全决定于 T^{n-1} 的元素,且这个矩阵容易算出. 在现在的情况

$$p_{55}^{(n)} = p_{56}^{(n)} = \frac{1}{4}\left(\frac{5}{12}\right)^{n-1} \text{ 因而 } \lambda_n = \frac{7}{48}\left(\frac{5}{12}\right)^{n-2}.$$

(b) **兄妹交配.** 作为第二个例子我们给出第十五章例 (2.1) 的详细讨论. 容易看出, E_1 与 E_5 各构成一个闭集(从生物学上的意义来看这是一个明显的事实). 如果系统从任一个其它状态出发, 则它最终将进入 E_1 或 E_2 并停留在那里. 饲养员想要知道对应的概率及过程持续时间的期望值.

删去第一、五两行与第一、五两列,我们得到简化后的矩阵

$$T = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{8} \\ 0 & \frac{1}{4} & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}. \tag{4.7}$$

现在可用第 1 节中的方法来计算幂 T^n. 它们表示非常返状态中的转移概率.

方程 (1.7a) 化归为

$$x_1 = \frac{s(2x_1 + x_2)}{4}, \quad x_2 = \frac{s(2x_1 + 2x_2 + 2x_3 + x_4)}{8},$$

$$x_3 = \frac{s(x_2 + 2x_3)}{4}, \quad x_4 = sx_2. \tag{4.8}$$

这组方程仅当行列式为零时才有解, 而这个条件则导致关于 s 的一个四次方程. 为书写简单起见,我们令

$$\theta_1 = 5^{\frac{1}{2}} - 1, \quad \theta_2 = 5^{\frac{1}{2}} + 1. \tag{4.9}$$

于是四个根 s_r 为

$$s_1 = 2, \quad s_2 = 4, \quad s_3 = \theta, \quad s_4 = -\theta_2, \tag{4.10}$$

而(4.8)的对应解 $(x_1^{(r)}, \cdots, x_4^{(r)})$ 为

$$\begin{gathered}(1, 0, -1, 0), (1, -1, 1, -4), \\ (1, \theta_1, 1, \theta_1^2), \quad (1, -\theta_2, 1, \theta_2^2).\end{gathered} \tag{4.11}$$

关于 $y_k^{(r)}$ 的线性方程组可通过特殊化而得到. 由(1.7b), 四组解为

$$\begin{gathered}(1, 0, -1, 0), \quad (1, -1, 1, -1/2), \\ (1, \theta_1, 1, \theta_1^2/8), \quad (1, -\theta_2, 1, \theta_2^2/8).\end{gathered} \tag{4.12}$$

由(1.11)我们求得四个常数 $c_1 = 1/2$, $c_2 = 1/5$, $c_3 = \theta_2^2/40$, $c_4 = \theta_1^2/40$. 由(1.8)我们得 $\rho_{jk}^{(r)}$; 最后, 对于所有的非常返状态, 即对于 $j, k = 2, 3, 4, 6$, $p_{jk}^{(n)}$ 由(1.3)给出. 对于固定的 j, k, 序列 $p_{jk}^{(n)}$ 是公比为 $s_1, \cdots, s_4$ 的四个几何级数的和.

仅当第 $n-1$ 使系统处于 E_2 或 E_3 而第 n 步使系统处于 E_1 时,才有可能恰好于第 n 步在 E_1 吸收. 这个事件的概率为 $p_{j2}^{(n-1)}/4 + p_{j3}^{(n-1)}/16$. 类似地, 在 E_5 处吸收的概率为 $p_{j3}^{(n-1)}/16 + p_{j4}^{(n-1)}/4$. 对所有的 n 求和,我们分别得到系统最终将到达并停留在 E_1 与 E_2 的概率. 这些概率的实际计算仅需对四个几何级数求和.

5. 在循环时间中的应用

在第十三章问题 19 中曾指出, 循环事件 $\mathscr{E}$ 的循环时间的均值 μ 与方差 σ^2 能用 $\mathscr{E}$ 在第 n 次试验出现的概率 u_n 来计算. 如果 $\mathscr{E}$ 不是周期的,且 σ^2 有限,则

$$u_n \to \frac{1}{\mu} \quad 且 \quad \sum_{n=0}^{\infty}\left(u_n - \frac{1}{\mu}\right) = \frac{\sigma^2 - \mu + \mu^2}{2\mu^2}. \tag{5.1}$$

如果我们把一个常返状态 E_j 看作是 $\mathscr{E}$,则 $u_n = p_{jj}^{(n)} (u_0 = 1)$. 因为在有限马尔科夫链中所有循环时间具有有限的方差(参看第十五章问题 19), 故(5.1)可以应用. 设 E_j 不是周期的且公式

(1.3)成立．于是 $s_1 = 1$ 且 $|s_r| > 1\ (r = 2, 3, \cdots)$，故 $p_{jj}^{(n)} \to \rho_{jj}^{(1)} = 1/\mu_j$．公式

$$p_{jj}^{(n)} - \frac{1}{\mu_j} = \sum_{r=2}^{a} \rho_{jj}^{(r)} s_r^{-n} \tag{5.2}$$

与(5.1)中的项 $u_n - 1/\mu$ 相对应．这个公式当 $n \geqslant 1$ 时成立；求公比为 s_r^{-1} 的几何级数的和，我们求得

$$\sum_{n=1}^{\infty} \left(p_{jj}^{(n)} - \frac{1}{\mu_j} \right) = \sum_{r=2}^{a} \frac{\rho_{jj}^{(r)}}{s_r - 1}. \tag{5.3}$$

将此式引入(5.1)，我们得到：如果 E_j 是非周期常返状态，则其平均循环时间由 $\mu_j = 1/\rho_{jj}^{(1)}$ 给出，循环时间的方差则为

$$\sigma_j^2 = \mu_j - \mu_j^2 + 2\mu_j^2 \sum_{r=2}^{a} \frac{\rho_{jj}^{(r)}}{s_r - 1}. \tag{5.4}$$

当然，我们假定公式(1.3)可以应用且 $s_1 = 1$．在周期状态的情况及出现重根时，只需作一些显然的修改．

第十七章 随机过程初步[1]

1. 一般讨论

随机徘徊与马尔科夫链是仅在固定时刻，比方说 $t=1,2,3,\cdots$ 发生改变的随机过程. 另一方面，在第六章第5，6两节中，我们已涉及到象电话呼唤、放射性蜕变和染色体的分裂之类的现象，其中变化可在任何时刻发生. 显然，这类过程的完全描述超出离散概率的范围. 为确定起见，考虑一个电话交换台收到的呼唤（倒不如说是实际过程的一个理想化的数学模型）. 每个时刻 t 对应于一个试验，一次实验的结果可用给出直到时刻 t 的呼唤数的一个函数 $\mathbf{X}(t)$ 来描述. 如果第一次呼唤在时刻 t_1 出现，第二个在 t_2，等等，则函数 $\mathbf{X}(t)$ 当 $0<t<t_1$ 时等于 0，当 $t_1<t<t_2$ 时等于 1，当 $t_2<t<t_3$ 时等于 2，等等. 反之，每个仅取值 $0,1,2,\cdots$ 的不减函数 $\mathbf{X}(t)$ 表示在我们的电话交换台处的一种进程. 换句话说，我们的理想实验的完全描述需要以函数 $\mathbf{X}(t)$ 为样本点的一个样本空间（在离散试验的情况样本点为一序列）. 象"在某天的一分钟内有 7 次呼唤"这样的一个复合事件显然是满足如下条件的 $\mathbf{X}(t)$ 的一个集合：对于指定区间中的某点 t 有 $\mathbf{X}(t+h)-\mathbf{X}(t)\geqslant 7$，其中 h 表示一分钟的长度.

这里我们不能考虑这样复杂的样本空间，而必须推迟对理论的更精的方面的研究. 幸运的是，对于某些有趣的问题，用简单的方法就能加以回答.

如果我们限制考虑任意固定的长为 t 的一段时间内的呼唤数 $\mathbf{X}(t)$，则 $\mathbf{X}(t)$ 是取值 $0,1,2,\cdots$ 的一个具有熟悉类型的随机变量.

---

1) 本章几乎与第十一—十六章无关.

设 $P_n(t)$ 为 $\mathbf{X}(t)=n$ 的概率. 诚然这时分布 $\{P_n(t)\}$ 依赖于一个连续的参数,但本书所引入的大多数分布都是这样.

这种情形最好用普阿松分布

$$P_n(t)=e^{-\lambda t}\frac{(\lambda t)^n}{n!} \tag{1.1}$$

来举例说明. 它曾在第六章第5节中作为二项分布的极限分布而导出; 在第十二章第3节中包含有一个更令人满意的推导. 我们将不利用那章的结果,但那里所分析的情况是如此简单和典型,所以一个简短的总结可作为本章的一个最好的导言.

考虑用整值随机变量 $\mathbf{X}(t)\geqslant 0$ 所表示的一个随机过程. 直观上我们可将 $\mathbf{X}(t)$ 解释为, 比方说, 雷电所造成的以元计的累积损失. 如果引入如下两个简单的假设, 则可得出一个特别简单的数学模型. 从0到 $t+s$ 的时间间隔内的增量 $\mathbf{X}(t+s)-\mathbf{X}(0)$ 是对应于从0到 s 与从 s 到 $t+s$ 的子间隔的增量 $\mathbf{X}(s)-\mathbf{X}(0)$ 与 $\mathbf{X}(t+s)-\mathbf{X}(s)$ 之和. 首先我们假定这些增量是统计独立的,其次假定 $\mathbf{X}(t+s)-\mathbf{X}(s)$ 的分布仅依赖于 t(即仅依赖于间隔的长度,而不依赖于它的位置: 这是对时间的均匀性).

设 $h_n(t)$ 为 $\mathbf{X}(t+s)-\mathbf{X}(s)$ 取值 n(此处 $n=0,1,2,\cdots$)的概率. 解析上, $\mathbf{X}(t+s)-\mathbf{X}(s)$ 与 $\mathbf{X}(s)-\mathbf{X}(0)$ 的独立性用

$$h_n(t+s)=\sum_{j=0}^{n}h_j(s)\cdot h_{n-j}(t) \tag{1.2}$$

来表示. 在第十二章第3节中已经证明,具有性质(1.2)的唯一分布 $\{h_n(t)\}$ 是复合普阿松分布;即 $\mathbf{X}(t)$ 具有随机变量

$$\mathbf{S_N}\left(\text{其中 } \mathbf{P}\{\mathbf{N}=n\}=e^{-\lambda}\frac{(\lambda t)^n}{n!}\right) \tag{1.3}$$

的分布, 此处 $\mathbf{S}_n=\mathbf{Y}_1+\mathbf{Y}_2+\cdots+\mathbf{Y}_n$ 为 n 个具有共同分布 $\{f_i\}$ $(i=0,1,2,\cdots)$ 的独立随机变量之和. 在我们的例子中, $\{f_i\}$ 表示个别雷击所造成的损失的概率分布; (1.3) 表明, 在长为 t 的时间间隔中的击中数服从普阿松分布 (1.1), 且各次损失是相互独立的随机变量. 变量 (1.3) 与任意长为 t 的时间间隔内的

变化 $\mathbf{X}(t+s)-\mathbf{X}(s)$ 具有相同的分布，又我们看到，这个总变化是随机个（$\mathbf{N}$ 个）个别变化或跳跃之和．变化的数目 $\mathbf{N}$ 具有普阿松分布（1.1），而个别的跳跃具有概率分布 $\{f_i\}$．特别，普阿松分布（1.1）本身表示所有的跳跃具有单位长度（即 $f_1=1$，$f_0=f_2=\cdots=0$，变量 $\mathbf{Y}_n$ 仅取值 1）的特殊情况．

读者一定注意到，我们已通过内在的概率性质求得简单和复合普阿松分布的一个特征．普阿松分布不再作为其它分布的一种近似或极限形式出现，而是根据它自身的性质而处于独立的地位（或者可以说，它表达了一种物理定律）．它的推导具有纯粹的分析特征，随机过程的概念和随机变量 $\mathbf{X}(t)$ 仅用来得到关于分布 $\{h_n(t)\}$ 的一组看来合理的假设．在很多应用中仅需要关于 $\{h_n(t)\}$ 的知识．从理论上说，应当证明，$\{h_n(t)\}$ 确实决定一类随机变量 $\mathbf{X}(t)$ 以及所有象事件"$\mathbf{X}(t)$ 总要超过 $at+b$"的概率（这是保险中的集体风险理论的输光问题）那样的有关概率关系．

这种类型的问题超出本书的范围．我们将满足于将过程的物理描述解释为只依赖于基本概率 $P_n(t)$ 的某些性质，并把 $P_n(t)$ 看作是依赖于 t 的一类离散概率分布．

这种局限于离散概率的人为限制，有不可避免的缺欠．例如，考虑（1.1）中的零项．我们将

$$P_0(t)=e^{-\lambda t} \tag{1.4}$$

解释为在长为 t 的一段观察时间内没有呼唤出现的概率．这个说法显示出，$P_0(t)$ 可以解释为直到第一次呼唤的等待时间（从一个任意的时刻开始）超过 t 的概率．可以证明，这个解释是正确的，但要注意，它牵涉一个连续统的概率．我们的第一种说明的操作意义如下．作具有固定观察时间 t 的一系列"恒等的观察"．每次试验的结果为"无呼唤"（成功）或"有一个或多个呼唤"（失败）．于是我们得到成功概率为 $e^{-\lambda t}$ 的伯努利试验．第二种解释是，我们一直等待到一个呼唤的来到．每个正数都是一个可能的等待时间，所以对应于每次试验的样本空间是半直线 $t>0$．于是公式（1.4）表示一个连续的概率分布．

2. 普阿松过程

首先我们给出普阿松分布一种新的推导；它决不比前面所叙述的推导更好，但它对我们将要研究的各种推广更适合.

考虑由于象物理质点的裂变、电话呼唤的到来或在有害辐射下染色体的分裂之类的随机事件的出现而发生变化的系统. 假定所有的变化具有同一类型，且我们仅考虑它们的总数. 每一变化用时间轴上的一个点来表示，所以我们是在研究直线上点的某种随机分布.

我们所考虑的物理过程可用如下两个性质来刻划：它们在时间上是均匀的，且未来的变化与过去的变化无关. 这意味着，确定过程的力保持绝对不变，所以对于长为 t 的所有时间间隔，任何特殊事件的概率都相同，而不依赖于这个间隔的位置及系统过去的历史[1].

现在我们将这种说明翻译为数学语言. 过程可用在长为 t 的时间间隔中恰有 n 次变化出现的概率[2] $P_n(t)$ 来描述. 特别，$P_0(t)$ 是无变化的概率，$1-P_0(t)$ 为有一个或多个变化的概率. 我们假定，当 $t\to 0$ 时有[3]

$$\frac{1-P_0(t)}{t}\to\lambda, \tag{2.1}$$

此处 λ 为常数. 于是，在长为 h 的小间隔中有一个或多个变化的概率为 $1-P_0(h)=\lambda h+o(h)$，其中 $o(h)$ 表示阶低于 h 的阶的

---

1) 在电话交换台中，在一天最繁忙的期间来到的呼唤比，譬如说，半夜12点与1点之间要更经常一些；因此过程在时间上是不均匀的. 不过，显然电话工程师所关心的主要是一天的"繁忙期间"，在这期间中可将过程看作是均匀的. 经验也表明，在繁忙期间中来到的流惊人精确地服从普阿松分布. 类似的考虑可应用于汽车事故，它们在星期天中更经常一些，等等.

2) 对于非均匀过程，我们必须引入在间隔 $t_1<t<t_2$ 中出现 n 个变化的概率 $P_n(t_1, t_2)$.

3) 这个假设可以省略，见第6节.

量．现在我们来确切陈述我们的

关于普阿松过程的假设． 不论在 $(0, t)$ 中变化的数目怎样，在 $(t, t+h)$ 中出现一个变化的概率为 $\lambda h + o(h)$，而且出现一个以上的变化的概率为 $o(h)$．

由这个条件容易导出关于 $P_n(t)$ 的一组微分方程．考虑两个邻接区间 $(0, t)$ 与 $(t, t+h)$，此处 h 很小．如果 $n \geqslant 1$，则在间隔 $(t, t+h)$ 出现 n 个变化恰有如下三种互不相容的方式：(1) 在 $(t, t+h)$ 中无变化，而在 $(0, t)$ 中有 n 个变化；(2) 在 $(t, t+h)$ 中有一个变化，而在 $(0, t)$ 中有 $n-1$ 个变化；(3) 在 $(t, t+h)$ 中有 $x \geqslant 2$ 个变化，而在 $(0, t)$ 中有 $n-x$ 个变化．按照我们的假设，第一个事件的概率为 $P_n(t)$ 与 $(t, t+h)$ 中无变化的概率（此概率为 $1-\lambda h-o(h)$）的乘积．类似，第二个事件具有概率 $P_{n-1}(t)\lambda h + o(h)$，而最后一个事件的概率则是其阶低于 h 的阶的量．于是我们有

$$P_n(t+h) = P_n(t)(1-\lambda h) + P_{n-1}(t)\lambda h + o(h) \tag{2.2}$$

或

$$\frac{P_n(t+h) - P_n(t)}{h} = -\lambda P_n(t) + \lambda P_{n-1}(t) + \frac{o(h)}{h}. \tag{2.3}$$

当 $h \to 0$ 时，最后一项趋向于零，所以左边的极限[1]存在，且

$$P'_n(t) = -\lambda P_n(t) + \lambda P_{n-1}(t) \quad (n \geqslant 1).$$

对于 $n = 0$，以上提到的第二个和第三个事件不会出现，故 (2.4) 被如下的较简单的方程来代替：

$$P_0(t+h) = P_0(t)(1-\lambda h) + o(h), \tag{2.5}$$

由此得

$$P'_0(t) = -\lambda P_0(t). \tag{2.6}$$

由 (2.6) 及 $P_0(0) = 1$ 我们得 $P_0(t) = e^{-\lambda t}$．以这个 $P_0(t)$ 代

---

1) 因为我们限制 h 取正值，故 (2.4) 中的 $P'_n(t)$ 应解释为右导数．它实际上是通常的双侧导数．事实上，(2.2) 中的项 $o(h)$ 不依赖于 t，故当 $t-h$ 代替 t 时保持不变．于是 (2.2) 蕴涵连续性，而 (2.3) 内蕴涵通常意义下的可微性．这一附注适用于整个本章内容，以后不再重复．

入 (2.4) ($n=1$), 我们得到关于 $P_1(t)$ 的一个微分方程. 因为 $P_1(0)=0$, 故我们容易求得 $P_1(t)=\lambda te^{-\lambda t}$, 这与普阿松分布(1.1)一致. 用同样的方法,我们依次可求得 (1.1) 中的所有项.

3. 纯增殖过程

在普阿松过程中, 在 $(t, t+h)$ 中出现一个变化的概率不依赖于 $(0, t)$ 中的变化数目. 放弃这个假设便得到一种最简单的推广. 代替这个条件,我们假定, 当 $(0, t)$ 出现 n 个变化时, 在 $(t, t+h)$ 中出现一个新变化的概率等于 $\lambda_n h$ 与阶低于 h 的阶的项之和;刻划过程的单一常数 λ 被序列 $\lambda_0, \lambda_1, \lambda_2, \cdots$ 所代替.

引入一个更灵活的术语是便利的. 如果在 $(0, t)$ 中出现 n 个变化,则我们说系统处于状态 E_n. 于是一个新的变化成为一个转移 $E_n\to E_{n+1}$. 在纯增殖过程中,由 E_n 仅能转移到 E_{n+1}. 这样一个过程由如下的假设来刻划:

假设. 如果在时刻 t 系统处于状态 $E_n(n=0, 1, 2, \cdots)$, 则在 $(t, t+h)$ 转移到 E_{n+1} 的概率等于 $\lambda_n h+o(h)$; 任何其它变化的概率为 $o(h)$.

这个假设的突出特征是, 系统在任何特殊状态所花费的时间不起作用;只要系统停留在单一状态,就会有状态的突然改变而不会有老化现象.

再次令 $P_n(t)$ 为在时刻 t 系统处于状态 E_n 的概率. 函数 $P_n(t)$ 满足一组微分方程,它能用上节中的方法导出,仅有的改变是用

$$P_n(t+h)=P_n(t)(1-\lambda_n h)+P_{n-1}(t)\lambda_{n-1}h+o(h) \quad (3.1)$$

来代替 (2.2). 用这种方法我们得到基本微分方程组

$$P'_n(t)=-\lambda_n P_n(t)+\lambda_{n-1}P_{n-1}(t) \quad (n\geqslant 1),$$
$$P'_0(t)=-\lambda_0 P_0(t). \quad (3.2)$$

我们能首先计算 $P_0(t)$, 然后依次求出所有的 $P_n(t)$. 如果系统的状态表示 $(0, t)$ 中变化的数目, 则初始状态为 E_0, 故 $P_0(0)=1$ 因而 $P_0(t)=e^{-\lambda_0 t}$. 不过, 系统不必从状态 E_0 出发[见例 (3.b)].

如果在时刻零系统处于 E_i，则我们有

$$P_i(0)=1, P_n(0)=0 \quad (n\neq i). \tag{3.3}$$

这些初始条件唯一地确定(3.2)的解 $\{P_n(t)\}$.（特别，$P_0(t)=P_1(t)=\cdots=P_{i-1}(t)=0$.）关于 $P_n(t)$ 的明显公式已被很多作者独立地导出，但对我们并无重要关系. 容易验证，对于任意指定的 λ_n，系统 $\{P_n(t)\}$ 具有全部所需性质，只是在某些条件下 $\sum P_n(t)<1$. 这个现象将在第 4 节中讨论.

例. (a) **放射性蜕变.** 放射性原子，例如铀原子，可以通过质点或 γ 射线的发射而变成另一种原子. 每种原子表示系统的一个可能状态，当过程继续下去时，我们便得到接连的转移 $E_0\to E_1\to E_2\to\cdots\to E_m$. 根据物理理论，只要原子处于状态 E_n，转移 $E_n\to E_{n+1}$ 的概率就保持不变，而这个假设是用我们的初始假定来表示的. 所以过程由微分方程 (3.2) 来描述(对于物理学家来说这是一个熟知的事实). 如果 E_m 是不能由它作进一步的转移的终了状态，则 $\lambda_m=0$ 且方程组(3.2)截止于 $n=m$.（对于 $n>m$ 我们自动地得到 $P_n(t)=0$.）

(b) **尤尔 (Yule) 过程.** 考虑一个总体，其成员能(通过分裂或其它方式)生出新成员而不致死亡. 在长为 h 的任何短的时间间隔中，每个成员产生一个新成员的概率为 $\lambda h+o(h)$；常数 λ 确定总体的增长速度. 如果成员之间无相互作用且在时刻 t 总体的大小为 n，则在 $(t, t+h)$ 中总体增大的概率为 $n\lambda h+o(h)$. 因此总体恰有 n 个元素的概率 $P_n(t)$ 满足 (3.2)(其中 $\lambda_n=n\lambda$)，即

$$P_n'(t)=-n\lambda P_n(t)+(n-1)\lambda P_{n-1}(t) \quad (n\geqslant 1). \tag{3.4}$$

如果 i 是在 $t=0$ 时总体的大小，则初始条件 (3.3) 成立. 容易验证，当 $n\geqslant i$ 时解由

$$P_n(t)=\binom{n-1}{n-i}e^{-i\lambda t}(1-e^{-\lambda t})^{n-i} \tag{3.5}$$

给出；显然，当 $n<i$ 时 $P_n(t)=0$. 这个分布是负伯努利分布的特殊情况：利用第六章中的定义 (8.1) 我们可将 (3.5) 改写为 $P_n(t)=f(n-i; i, e^{-\lambda t})$. 由此可知[参看第九章例 (3.c)]，在时

刻 t 总体的大小是 i 个独立随机变量之和，其中每个变量的分布可由在(3.5)中用1代替 i 而得. 这 i 个变量表示我们的总体 i 个原始成员的后裔.

这种类型的过程首先被尤尔[1]联系到数学进化论而研究. 总体由一个属中的物种组成,新元素的产生是由于突变.每一物种具有产生新种的相同概率的假设忽略了物种大小上的差别. 因为我们也忽略了物种灭绝的可能性,故可以预料，公式(3.5)仅给出一个粗糙的近似. 弗里[54] (W. H. Furry) 曾用相同的模型描述与宇宙射线有关的过程,但近似也是相当粗糙的. 微分方程(3.4)严格地适用于能分裂为它们自身的精确副本的质点的总体,当然,这以质点之间没有相互作用为条件.

*4. 发散的增殖过程

无限微分方程组(3.2)满足初始条件(3.3)的解可从 $P_i(t) = e^{-\lambda_i t}$ 出发归纳地算出. 所以分布 $\{P_n(t)\}$ 被唯一地确定. 由解线性微分方程的熟悉公式也可推出 $P_n(t) \geqslant 0$. 剩下的未解决的问题是，$\{P_n(t)\}$ 是否确实为一概率分布,即是否对所有的 t 有

$$\sum P_n(t) = 1. \tag{4.1}$$

我们将看到，情况并不总是如此：如果系数 λ_n 增加得充分快，则可能有

$$\sum P_n(t) < 1. \tag{4.2}$$

乍一看来,这种可能性似乎是奇怪和麻烦的,但它有如下一种简单的解释. (4.2)的左边可以解释为在时刻 t 仅发生有限个变化的

---

1) 见[50],该书以[51]中的结论为基础. 尤尔并未引入微分方程(3.4),而是用类似于第六章第5节关于普阿松过程所使用的极限过程来推导 $P_n(t)$. 姆'肯德里克 (A. G. M'Kendrick) 在[52]中设计了更一般、更灵巧的模型，并将它们应用于流行病和总体的增长中. 遗憾的是，这篇非凡的论文实际上并未受到注意. 特别是，本书作者在[53]中引入总体增长的各种随机过程模型时也不知道[52]中的结果.

*本节讨论专门的论题,可以略去.

概率．所以（4.2）两边的差正说明无限多个变化或一种爆炸的可能性．为了更好地理解这个现象，让我们将我们的增长概率模型与熟知的决定论方法相比较．

（3.2）中的量 λ_n 可称为总体大小为 n 时的平均增长速度．例如，在特殊情况（3.4）我们有 $\lambda_n = n\lambda$，故平均增长速度与总体的实际大小成比例．如果增长不受随机起伏的影响且增加的速度与瞬时的总体大小成比例，则 $x(t)$ 满足如下决定论的微分方程

$$\frac{dx(t)}{dt} = \lambda x(t). \tag{4.3}$$

由此可以推出，在时刻 t 总体的大小为

$$x(t) = ie^{\lambda t}, \tag{4.4}$$

此处 $i = x(0)$ 是初始的总体大小．（3.4）与（4.3）之间的联系并不纯粹是形式上的．容易看到，（4.4）实际上给出分布（3.5）的期望值，故（4.3）描述期望的总体大小，而（3.4）则考虑随机起伏．

现在让我们考虑一个决定论的增长过程，其中增长速度增加得比总体的大小快．当增长速度与 $x^2(t)$ 成比例时，对应的微分方程为

$$\frac{dx(t)}{dt} = \lambda x^2(t), \tag{4.5}$$

其解为

$$x(t) = \frac{i}{1 - \lambda it}. \tag{4.6}$$

注意，当 $t \to 1/\lambda i$ 时 $x(t)$ 无限增大．换句话说，增长速度按总体大小的平方增加的假设蕴涵在一个有限时间间隔中的无限增长．类似地，如果（3.4）中的 λ_n 增加得太快，则在一个有限的时间间隔中有无穷多个变化发生的概率为正，如下定理给出了使这样一个发散的增长出现的精确条件：

定理． 使（4.1）对所有的 t 成立的充分必要条件是级数

$$\sum \frac{1}{\lambda_n} \tag{4.7}$$

发散.

证. 令

$$S_k(t) = P_0(t) + \cdots + P_k(t), \tag{4.8}$$

我们由(3.2)得

$$S_k'(t) = -\lambda_k P_k(t), \tag{4.9}$$

故当 $k \geqslant i$ 时有

$$1 - S_k(t) = \lambda_k \int_0^t P_k(\tau)\, d\tau. \tag{4.10}$$

因为(4.8)中的所有项都是非负的，所以序列 $S_k(t)$——对于固定的 t——仅能随着 k 增加，故(4.10)的右边相对于 k 来说单调减少. 用 $\mu(t)$ 表示其极限. 于是对于 $k \geqslant i$ 有

$$\lambda_k \int_0^t P_k(\tau) d\tau \geqslant \mu(t), \tag{4.11}$$

故

$$\int_0^t S_n(\tau) d\tau \geqslant \mu(t) \left(\frac{1}{\lambda_i} + \frac{1}{\lambda_{i+1}} + \cdots + \frac{1}{\lambda_n} \right). \tag{4.12}$$

由于(4.10)，我们有 $S_n(t) \leqslant 1$，故(4.12)的左边最多为 t. 如果级数(4.7)发散，则(4.12)右边的第二个因子趋向于无穷，故对于所有的 t 仅当 $\mu(t) = 0$ 时不等式才能成立. 在这种情况，当 $k \to \infty$ 时(4.10)的右边趋向于零，因此 $S_n(t) \to 1$，故(4.1)成立. 反之，积分(4.8)并利用(4.10)，我们看到(4.12)的左边小于 $\lambda_0^{-1} + \lambda_1^{-1} + \cdots + \lambda_n^{-1}$. 如果级数(4.7)收敛，则这个表达式是有界的，因而对于所有的 t，$S_n(t) \to 1$ 是不可能的.

5. 生 灭 过 程

第3节中的纯增殖过程提供了放射性蜕变的一个令人满意的描述，但它不能作为其成员会死亡(或退出)的总体的大小的变化的现实模型. 这启发我们将这个模型作如下推广，不仅允许从状态 E_n 转移到邻近的较高的状态 E_{n+1}，也允许从 E_n 转移到邻近的较低的状态 E_{n-1}. (更一般的过程将在第9节中定义.) 因此我们

的出发点是如下的

假设. 系统的变化仅通过状态向其邻近状态转移（当 $n \geqslant 1$ 时由 E_n 到 E_{n+1} 或 E_{n-1}，但由 E_0 仅能到 E_1）而发生. 如果在任何时刻 t 系统处于状态 E_n，则在 $(t, t+h)$ 中出现转移 $E_n \to E_{n+1}$ 的概率等于 $\lambda_n h + o(h)$，而 $E_n \to E_{n-1}$（如果 $n \geqslant 1$）的概率等于 $\mu_n h + o(h)$. 在 $(t, t+h)$ 中出现一个以上的变化的概率为 $o(h)$.

容易修改第 2 节中的方法来导出关于系统在时刻 t 位于状态 E_n 的概率 $P_n(t)$ 的微分方程. 为计算 $P_n(t+h)$，我们注意，在时刻 $t+h$，仅当下列条件之一满足时，系统才可能处于状态 E_n：(1) 在时刻 t 系统处于 E_n 且在 $(t, t+h)$ 中无变化出现；(2) 在时刻 t 系统处于 E_{n-1} 且在 $(t, t+h)$ 中出现到 E_n 的一个转移；(3) 在时刻 t 系统处于 E_{n+1} 且在 $(t, t+h)$ 中出现到 E_n 的一个转移；(4) 在 $(t, t+h)$ 中出现两个或更多的转移. 根据假设，最后的事件的概率为 $o(h)$. 前三个事件是互不相容的，故它们的概率相加. 所以

$$\begin{aligned} P_n(t+h) = {} & P_n(t)\{1 - \lambda_n h - \mu_n h\} \\ & + \lambda_{n-1} h P_{n-1}(t) + \mu_{n+1} h P_{n+1}(t) + o(h). \end{aligned} \tag{5.1}$$

将项 $P_n(t)$ 移到左边并以 h 除等式两边，在左边我们得到 $P_n(t)$ 的差商. 令 $h \to 0$，我们得

$$P'_n(t) = -(\lambda_n + \mu_n)P_n(t) + \lambda_{n-1}P_{n-1}(t) + \mu_{n+1}P_{n+1}(t). \tag{5.2}$$

这个方程对于 $n \geqslant 1$ 成立. 对于 $n = 0$，用同样的方法可得

$$P'_0(t) = -\lambda_0 P_0(t) + \mu_1 P_1(t). \tag{5.3}$$

如果在时刻零系统处于状态 E_i，则初始条件为

$$P_i(0) = 1, \quad P_n(0) = 0 \quad (n \neq i). \tag{5.4}$$

于是我们看出，生灭过程依赖于无穷微分方程组(5.2)—(5.3)及初始条件 (5.4). 在这种情况，解的存在与唯一性问题决不是平凡的. 在纯增殖过程中，微分方程组 (3.2) 也是无穷的，但它有递推关系的形式；$P_0(t)$ 由第一个方程所决定，而 $P_n(t)$ 能由 $P_{n-1}(t)$ 来计算. 新方程组 (5.2) 并不具有这种形式，因而所有的 $P_n(t)$ 必

须同时求出．我们将在这里(以及本章其它地方)陈述解的性质而不给出其证明[1]．

对于任意指定的系数$\lambda_n \geqslant 0$，$\mu_n \geqslant 0$，总存在(5.2)—(5.4)的一组正解$\{P_n(t)\}$，使得$\sum P_n(t) \leqslant 1$．如果系数有界(或增加得充分慢)，则这个解是唯一的且满足正则性条件$\sum P_n(t) = 1$．不过，适当选择系数使得$\sum P_n(t) < 1$且存在无穷多个解也是可能的．在后一种情况，我们遇到在前一节中类似于纯增殖过程所研究过的现象．这种情况具有重要的理论意义[2]，但读者可以放心，在所有具有实际意义的情况，唯一性条件是满足的；在这种情况自动有$\sum P_n(t) = 1$(见第10节)．

当$\lambda_0 = 0$时，转移$E_0 \to E_1$是不可能的．用马尔科夫链的术语来说，E_0是一吸收状态，从它退出是不可能的；一但系统处于E_0，它就停留在那里．由(5.3)可以推出，在这种情况有$P_0'(t) \geqslant 0$，故$P_0(t)$单调增加．极限$P_0(\infty)$是最终吸收的概率．

更一般，可以证明，极限

$$\lim_{t\to\infty} P_n(t) = p_n \tag{5.5}$$

存在且不依赖于初始条件(5.4)；它们满足在(5.2)—(5.3)中令$P_n'(t) = 0$而得到的线性方程组．关系(5.5)通常解释为一种"朝向稳定状态条件的趋势"，而这个富于暗示的名称已引起了很多混乱．必须了解，除当E_0是吸收状态外，随机起伏仍然永远不会减少，(5.5)仅表明，初始条件的影响最终将会消失．第十五章第6节关于统计平衡的附注在这里完全适用．

(5.5)的正确性可由关于$P_n(t)$的明显公式或由一般的遍历性定理来证明．通过将我们的过程与具有转移概率

---

1) 适用于本章大多数一般方程的一个简单的存在证明和唯一性准则(虽然用到了拉氏变换)在[55]的第4节中被得到．第一个存在证明见[56]．遗憾的是，这篇论文处理非可数样本空间和系数依赖于时间的一般情况，且将其结果特殊化为本章所讨论的常系数常微分方程的情况就导致一个简单的存在证明这一事实，一般未受到注意．

2) 生灭过程的使得$\sum P_n(t) < 1$的解最近受到普遍的注意．见[57]与[58]．

$$p_{n,n+1}=\frac{\lambda_n}{\lambda_n+\mu_n},\quad p_{n,n-1}=\frac{\mu_n}{\lambda_n+\mu_n} \tag{5.6}$$

的简单马尔科夫链相比较，在直观上我们的定理变得几乎是显然的．在这个链中，仅有的直接转移为 $E_n\to E_{n+1}$ 与 $E_n\to E_{n-1}$，它们与我们的过程具有相同的条件概率；这个链与我们的过程之间的差别在于，对于后者，变化能在任意时刻出现，故在时间 t 内的转移数是一随机变量．不过，对于大的 t，这个数必然很大，因此当 $t\to\infty$ 时 $P_n(t)$ 的状态与简单链的对应条件概率相同表面上是讲得通的．

生灭过程主要在排队、占线等问题中有着应用；见第 6 节与第 7 节．

例．(a) **线性增长．**设总体由能分裂或死亡的元素组成．在任何长为 h 的短的时间间隔内一个活着的元素分裂为二的概率为 $\lambda h+o(h)$，而对应的死亡概率为 $\mu h+o(h)$．此处 λ 与 μ 是总体的两个常数特征．如果元素间无相互作用，则我们得到 $\lambda_n=n\lambda$，$\mu_n=n\mu$ 的生灭过程．基本微分方程具有如下形式：

$$\begin{aligned}P_0'(t)&=\mu P_1(t),\\ P_n'(t)&=-(\lambda+\mu)nP_n(t)+\lambda(n-1)P_{n-1}(t)\\ &\quad+\mu(n+1)P_{n+1}(t).\end{aligned} \tag{5.7}$$

显解可以求出[1)]（参看问题 9—11），但此处不打算讨论这方面的问题．极限 (5.5) 存在且满足具有 $P_n'(t)=0$ 的方程 (5.7)．由第一个方程我们得到 $p_1=0$，又由归纳法从第二个方程可以看出，对于所有的 $n\geqslant1$ 有 $p_n=0$．如果 $p_0=1$，则我们可以说，最终绝灭的概率为 1．如果 $p_0<1$，则关系 $p_1=p_2=\cdots=0$ 蕴涵总体以概率 $1-p_0$ 增大超过任何界限；最后总体必定逐渐消灭或无限增加．为了求得绝灭概率 p_0，我们将过程与有关马尔科夫链相比较．在我们的情况，转移概率 (5.6) 不依赖于 n，所以我们有一

1）通常的解法是，导出关于母函数 $\sum P_n(t)s^n$ 的偏微分方程．在 [59] 中详细讨论了允许 (5.7) 中的系数依赖于时间的一个更一般的过程．也见同一作者的 [60]，在那里将理论作了推广，以考虑生物总体的年龄分布．

个通常的随机徘徊，其中向右移动一步与向左移动一步的概率分别为 $p=\lambda/(\lambda+\mu)$ 与 $q=\mu/(\lambda+\mu)$. 状态 E_0（或 $x=0$）是一个吸收壁. 由古典输光问题可知(见第十四章第 2 节)，如果 $p\leqslant q$，则绝灭概率为 1；如果 $q<p$，则这个概率为 $(q/p)^r$，其中 r 为初始状态. 我们得到如下结论：在我们的过程中，最终绝灭的概率，当 $\lambda\leqslant\mu$ 时为 1，当 $\lambda>\mu$ 时为 $(\mu/\lambda)^r$.（这容易由显解来验证；见问题 10.）

像在很多类似的情况中那样，(5.7) 的显解是相当复杂的，因此我们希望直接从微分方程来计算分布 $\{P_n(t)\}$ 的均值与方差. 对于均值我们有

$$M(t)=\sum_{n=1}^{\infty} nP_n(t). \tag{5.8}$$

我们将略去 $M(t)$ 为有限以及下面的形式运算为合理（这两者均容易由问题 10 中所给的解推出）的正式证明. 以 n 乘 (5.7) 中的第二个方程，并对 $n=1, 2, \cdots$, 求和，我们发现含有 n^2 的项可以消去，于是我们得

$$\begin{aligned} M'(t) &= \lambda\sum(n-1)P_{n-1}(t)-\mu\sum(n+1)P_{n+1}(t) \\ &= (\lambda-\mu)M(t). \end{aligned} \tag{5.9}$$

这是关于 $M(t)$ 的一个微分方程. 在时刻 $t=0$ 总体大小为 i，故 $M(0)=i$. 所以

$$M(t)=ie^{(\lambda-\mu)t}. \tag{5.10}$$

我们看到，均值趋向于 0 或无穷，按照 $\lambda<\mu$ 或 $\lambda>\mu$ 而定. $\{P_n(t)$ 的方差能用类似的方法来计算(参看问题 12).

(b) **单线路的排队问题**. 在常系数 $\lambda_n=\lambda$, $\mu_n=\mu$ 的最简单情况，生灭过程化归为例 (7.b) 中的排队问题当 $a=1$ 的特殊情况.

6. 指数持续时间

生灭过程的主要应用领域与电话工程中的占线以及电话、计

数器或机器中的各种排队问题有关．这种类型的问题能够用近似程度不同的各种数学模型来处理．生灭过程的方法提供一种最容易的门径，但这个模型以指数持续时间假设这一数学简化为基础．我们首先讨论这个基本假设．

为语言具体起见，让我们首先考虑电话会话，我们假定，会话的长度必须为秒的整倍数．我们把会话的长度作为一个随机变量来处理，并假定其概率分布 $p_n = \mathbf{P}\{\mathbf{X} = n\}$ 为已知．于是电话线路表示一个物理系统，它的两个状态是“忙碌”(E_0)和“空闲”(E_1)．如果在一个任意的时刻 t 线路是忙碌的，则在下一秒中状态发生变化的概率与会话已进行多久有关．换句话说，过去对将来有影响，因此我们的过程不是马尔科夫过程(见第十五章第10节)．在更复杂的问题中，这种情况是大多数困难的根源．不过，存在一种我们曾在第十三章第9节中详细讨论过的简单的例外情况．

设想在每一秒钟会话是否继续下去，由一个不对称的钱币来随机地决定．换句话说，一列成功概率为 p 的伯努利试验以每秒一次的速度进行，直到出现第一次成功为止．当这个第一次成功出现时会话结束．在这种情况，会话的总长，即“持续时间”，具有几何分布 $p_n = q^{n-1}p$．如果在任何时刻 t 线路是忙碌的，则它在下一秒仍然忙碌的概率为 q，而出现转移 $E_0 \to E_1$ 的概率则为 p．在现在的情况这些概率与线路忙碌有多久无关．

如不将时间参数离散化，则我们必须处理连续型随机变量．这时等待时间的几何分布的作用由指数分布来接替．这是具有马尔科夫性亦即完全缺乏记忆的仅有分布．换句话说，在时刻 x 进行的会话继续进行到 $x+h$ 以后的概率与会话已持续的时间无关，其充要条件是，会话持续 t 个时间单位以上的概率由指数函数 $e^{-\lambda t}$ 给出．这个“指数持续时间分布”就是普阿松分布(1.1)中的零项，亦即直到第一次变化出现的等待时间超过 t 的概率．

生灭过程的方法仅当问题中的转移概率不依赖于过去时才能应用；对于占线和排队问题，这意味着所有持续时间必须是指数的．从实际观点来看，这个假设乍一看来可能显得有点不自然，但

经验表明，它合理地描述了实际现象．特别，很多测量都证明，在一个城市中的电话会话[1]惊人精确地服从指数律．对于其它的持续时间（例如，机器修理的持续时间），情况也是如此．

尚须刻划所谓来到的流（来到的呼唤、机器的故障等）．我们假定，在任何长为 h 的时间间隔中来到一个呼唤的概率为 λh 与一个可忽略项之和，且多于一个呼唤的概率在取极限时可以略去．根据第 2 节的结果，这意味着来到的呼唤具有均值为 λt 的普阿松分布．在这种情况我们说来到的流是强度为 λ 的普阿松型流．

容易验证指数持续时间的所述性质．用 $u(t)$ 表示会话至少持续 t 个时间单位的概率．从时刻 0 开始的会话持续到 $t+s$ 以后的概率 $u(t+s)$ 等于它持续 t 单位以上的概率乘已知会话长超过 t 的条件下，会话再持续 s 单位以上的条件概率．如果过去的持续时间无影响，则上述条件概率必须为 $u(s)$；即我们必须有

$$u(t+s)=u(t)u(s). \tag{6.1}$$

尚须证明的是

定理． 设 $u(t)$ 对于 $t>0$ 有定义，且在每个有限区间中有界．如果 $u(t)$ 满足 (6.1)，则或者对于所有的 t 有 $u(t)=0$，或者存在某个常数 λ，使得 $u(t)=e^{-\lambda t}$．

证．如果 $u(t)$ 不恒为零，则存在一个点 x，使得 $u(x)>0$．令 $\lambda=-\log u(x)$ 及 $v(t)=e^{\lambda t}u(xt)$．则

$$v(t+s)=v(t)v(s), \quad v(1)=1 \tag{6.2}$$

且我们将证明对于所有的 $t>0$ 有 $v(t)=1$．显然 $v^2(1/2)=v(1)=1$，一般对于每个整数 $n>0$ 有 $v^n(1/n)=v(1)=1$．所以 $v(1/n)=1$，由此对于每对整数 $m>0$，$n>0$ 有 $v(m/n)=v^m(1/n)=1$．故对于每个有理数 r 有 $v(r)=1$．现在设 $v(\tau)=c\neq 1$．于是 $v(\tau^{-1})=c^{-1}$，故我们可以假定 $c>1$．在这种情况，通过选取 N 充分大，就可使 $v(N\tau)=v^N(\tau)=c^N$ 任意大．现在我们在区间

$$N\tau-1<r<N\tau$$

中选取一个有理数．于是

$$v(N\tau-r)=v(N\tau-r)v(r)=v(N\tau)=c^N, \tag{6.3}$$

---

1) 对于城市间的会话，公司通常按三分钟的间隔收费．因此持续时间多半是三分钟的倍数．这与指数律有系统的偏差，于是我们的理论不能应用．

这表明在区间 $0 < a < 1$ 中存在一个点 $a = N\tau - r$，使得 $v(a) > c^N$. 这与 $u(t)$ 因而 $v(t)$ 在每个有限区间中有界的假设矛盾.

7. 排队与服务问题

(a) **最简单的占线问题**[1]. 设有无穷多条线路可以采用，且会话在间隔 $(t, t+h)$ 结束的概率为 μh 与当 $h \to 0$ 时可以略去的项之和(指数持续时间). 来到的呼唤构成强度为 λ 的普阿松型流. 如果有 n 条线路是忙碌的，则系统处于状态 E_n.

当然我们假定，会话的持续时间相互独立. 如果有 n 条线路是忙碌的，则在长为 h 的时间内它们之中将有一个空闲的概率为 $n\mu h + o(h)$. 在这段时间内有两个或多个会话结束的概率显然与 h^2 有相同的阶，因而可以略去. 来到一个新呼唤的概率为 $\lambda h + o(h)$. 几个呼唤的一个混合，或一个来到的呼唤与一个结束的会话的概率也是 $o(h)$. 于是由第 5 节中的记号，有

$$\lambda_n = \lambda, \quad \mu_n = n\mu. \tag{7.1}$$

基本微分方程 (5.2)—(5.3) 具有如下形式 ($n \geqslant 1$)

$$P_0'(t) = -\lambda P_0(t) + \mu P_1(t), \tag{7.2}$$

$$P_n'(t) = -(\lambda + n\mu)P_n(t) + \lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t).$$

显解可以通过导出关于母函数的一个偏微分方程而得到(参看问题 13). 我们只要确定 (5.5) 中的量 $p_n = \lim P_n(t)$. 它们满足方程

$$\lambda p_0 = \mu p_1,$$

$$(\lambda + n\mu)p_n = \lambda p_{n-1} + (n+1)\mu p_{n+1}. \tag{7.3}$$

用归纳法我们求得 $p_n = p_0(\lambda/\mu)^n/n!$，因此

---

1) 见 [61]. 关于电话交换台的排队和占线问题，早在随机过程的理论可以利用并对理论的发展产生激励的影响之前，就已被研究. 尤其是，已经表明，巴尔姆的给人深刻印象的工作，在其后很多年对好几个作者都有过帮助. 在这个领域中最早的工作者是埃尔兰 (A. K. Erlang) (1878—1929). 见 [62]. 具有独立价值的开拓性工作已由弗里完成，见 [63]；这本书为概率论的工程应用的发展做了许多工作.

$$p_n = e^{-\lambda/\mu}\frac{(\lambda/\mu)^n}{n!}. \tag{7.4}$$

于是，极限分布为具有参数 λ/μ 的普阿松分布．它与初始状态无关．

容易求出均值 $M(t) = \Sigma nP_n(t)$．以 n 乘(7.2)中的第 n 个方程并相加，考虑到 $P_n(t)$ 之和为1，我们得到

$$M'(t) = \lambda - \mu M(t). \tag{7.5}$$

如果初始状态为 E_i，则 $M(0) = i$，且

$$M(t) = \frac{\lambda}{\mu}(1 - e^{-\mu t}) + ie^{-\mu t}. \tag{7.6}$$

当 $t \to \infty$ 时，我们看到，$M(t)$ 趋向于上面所求得的普阿松分布的均值．读者可以验证，在 $i = 0$ 的特殊情况，$P_n(t)$ 精确地由均值为 $M(t)$ 的普阿松分布给出．

(b) **有限条线路的排队问题**[64]． 现在我们修改上面的例子，以得出一个更现实的模型．除线路数目有限外，假设与上例相同．如果所有的线路都是忙碌的，则每个新来的呼唤参加排队并一直等到有一条线路空闲为止．这意味着所有的线路有一个公共的排队列．

"线路"一词可用邮局中的柜台来代替，而会话则用服务来代替．实际上我们所讨论的是仅当所有 a 条线路都忙碌时才需等待的一般排队问题．

如果恰有 n 个人在接受服务或在排队，则我们说系统处于状态 E_n．仅当 $n > a$ 时才出现排队，此时参加排队的共有 $n - a$ 人．

只要仅有一条线路空闲，情况就和前面的例子完全一样．不过，如果系统处于状态 E_n，其中 $n > a$，则仅有 a 个会话在进行，因而当 $n \geqslant a$ 时我们有 $\mu_n = a\mu$．因此基本微分方程组当 $n < a$ 时由(7.2)给出，而当 $n \geqslant a$ 时则由

$$P'_n(t) = -(\lambda + a\mu)P_n(t) + \lambda P_{n-1}(t) + a\mu P_{n+1}(t) \tag{7.7}$$

给出．

在单一线路的特殊情况 ($a = 1$)，这些方程化归为系数与 n

无关的生灭过程的方程.

(5.5) 的极限 p_n 存在，它们当 $n < a$ 时满足(7.3)，而当 $n \geqslant a$ 时则满足

$$(\lambda + a\mu)p_n = \lambda p_{n-1} + a\mu p_{n+1}. \tag{7.8}$$

由递推法可得

$$p_n = p_0 \frac{(\lambda/\mu)^n}{n!}, \qquad n \leqslant a, \tag{7.9}$$

$$p_n = \frac{(\lambda/\mu)^n}{a!\,a^{n-a}} p_0, \qquad n \geqslant a. \tag{7.10}$$

仅当

$$\frac{\lambda}{\mu} < a \tag{7.11}$$

时级数 $\sum(p_n/p_0)$ 收敛. 因此，如果(7.11)不成立，则极限分布 $\{p_k\}$ 不可能存在. 在这种情况对于所有的 n 有 $p_n = 0$，这意味着，排队的长度逐渐地无限增大. 另一方面，如果(7.11)成立，则我们能确定 p_0 使得(7.9)式与(7.11)式之和为 1. 虽然我们尚未导出 $P_n(t)$ 的显式，但由它可以证明，这样得到的 p_n 确实表示 $P_n(t)$ 的极限分布. 表 1 给出了 $a = 3$, $\lambda/\mu = 2$ 的一个数值例题.

表 1

当线路数 $a = 3$ 且 $\lambda/\mu = 2$时的极限概率

n	0	1	2	3	4	5	6	7
占线数	0	1	2	3	3	3	3	3
等待人数	0	0	0	0	1	2	3	4
p_n	0.1111	0.2222	0.2222	0.1481	0.0988	0.0658	0.0439	0.0293

(c) **机器维修问题**[1]. 本例与下例中所导出的结果已在瑞典工业中获得成功的应用. 为明确起见，我们首先讨论最简单的情况，然后在下一个例子中进行推广. 问题如下.

我们考虑在正常情况下不需人看管的自动机器. 但机器在任

1) 例(c)与(d)，包括数值例题，都取自 [65]. 最经济的修理工的数目的图表是由帕姆给出的.

何时刻都可能发生故障因而需要维修．机器所需的维修时间可以看作是一个具有指数分布的随机变量．换句话说，机器可用具有如下性质的两个常数 λ 与 μ 来刻划．如果在时刻 t 机器处于运转状态，则它在时刻 $t+h$ 之前需要维修的概率为 λh 与当 $h\to 0$ 时可以忽略的项之和．反之，如果在时刻 t 机器正在维修，则维修时间在 $t+h$ 之前结束因而机器恢复运转状态的概率为 $\mu h+o(h)$．对于高效率的机器 λ 应相对地小而 μ 则应相对地大．比值 λ/μ 称为维修因子．

我们假定，具有相同参数 λ 与 μ 且独立运转的 m 台机器由一个修理工负责维修．当机器发生故障时立即进行维修，除非修理工正在维修另一台机器，在后一种情况就形成排队．如果不运转的机器有 n 台，则我们说系统处于状态 E_n．$1\leqslant n\leqslant m$ 意味着有一台机器在维修，$n-1$ 台在排队（待修）；当系统处于状态 E_0 时，所有机器都运转而修理工则闲着．

转移 $E_n\to E_{n+1}$ 是由于 $m-n$ 台运转机器中有一台出了故障而引起的，而转移 $E_n\to E_{n-1}$ 则是由于有一台被维修的机器恢复运转状态而出现的．因此我们得到一个生灭过程，其系数为

$$\lambda_n=(m-n)\lambda,\ \mu_0=0,\ \mu_1=\mu_2=\cdots=\mu_m=\mu; \tag{7.12}$$

而基本微分方程（5.2）与（5.3）则变为（$1\leqslant n\leqslant m-1$）：

$$\begin{aligned} P_0'(t)&=-m\lambda P_0(t)+\mu P_1(t),\\ P_n'(t)&=-\{(m-n)\lambda+\mu\}P_n(t)\\ &\quad+(m-n+1)\lambda P_{n-1}(t)+\mu P_{n+1}(t),\\ P_m'(t)&=-\mu P_m(t)+\lambda P_{m-1}(t). \end{aligned} \tag{7.13}$$

这是一个有限的微分方程组，它可用通常的方法来解．极限（5.5）存在且满足方程

$$\begin{aligned} m\lambda p_0&=\mu p_1,\\ \{(m-n)\lambda+\mu\}p_n&=(m-n+1)\lambda p_{n-1}+\mu p_{n+1},\\ \mu p_m&=\lambda p_{m-1}. \end{aligned} \tag{7.14}$$

容易证明，如下的递推公式

$$(m-n)\lambda p_n=\mu p_{n+1} \tag{7.15}$$

成立. 依次以 $n = m-1, m-2, \cdots, 1, 0$ 代入,我们得

$$p_{m-k} = \frac{1}{k!}\left(\frac{\mu}{\lambda}\right)^k \cdot p_m. \tag{7.16}$$

未知常数 p_m 能由诸 p_i 之和为 1 的条件得到:

$$p_m = \left\{1 + \frac{1}{1!}\left(\frac{\mu}{\lambda}\right)^1 + \cdots + \frac{1}{m!}\left(\frac{\mu}{\lambda}\right)^m\right\}^{-1}. \tag{7.16a}$$

公式(7.16)是为线路工程师所熟知的埃尔兰损失公式. 在表 2 中列出了典型数值.

表 2

当 $\lambda/\mu = 0.1$, $m = 6$ 时的概率

(埃尔兰损失公式)

n	等待维修的机器数	p_n
0	0	0.4845
1	0	0.2907
2	1	0.1454
3	2	0.0582
4	3	0.0175
5	4	0.0035
6	5	0.0003

概率 p_0 可以解释为修理工闲着的概率(在表 2 的例子中他应当约有一半的时间是闲着的). 等待维修的机器的期望数为

$$w = \sum_{k=1}^{m}(k-1)p_k = \sum_{k=1}^{m} kp_k - (1-p_0). \tag{7.17}$$

这个量可以通过将关系式(7.15)对 $n = 0, 1, \cdots, m$ 求和来计算. 利用诸 p_n 之和为 1 这一事实,我们得到

$$m\lambda - \lambda w - \lambda(1-p_0) = \mu(1-p_0)$$

或

$$w = m - \frac{\lambda+\mu}{\lambda}(1-p_0). \tag{7.18}$$

在表 2 的例中我们有 $w = 6\cdot(0.0549)$. 于是 0.0549 是一台机器

对于排队列的平均贡献.

(d) 续上：**多个修理工的情况.** 我们将不改变上个问题中的基本假设，只是 m 台机器现在是由 r 个修理工 $(r < m)$ 来维修. 于是当 $n \leqslant r$ 时状态 E_n 意味着 $r - n$ 个修理工闲着，n 台机器在维修，而没有机器等待维修. 当 $n > r$ 时状态 E_n 意指，r 台机器在维修而 $n - r$ 台机器则在待修. 除了 (7.12) 显然应由

$$\begin{aligned} &\lambda_0 = m\lambda, \quad \mu_0 = 0, \\ &\lambda_n = (m-n)\lambda, \quad \mu_n = n\mu \quad (1 \leqslant n \leqslant r), \\ &\lambda_n = (m-n)\lambda, \quad \mu_n = r\mu \quad (r \leqslant n \leqslant m) \end{aligned} \tag{7.19}$$

来代替外，上例中的方法仍然适用. 我们将不写出基本微分方程组，而仅写出极限概率 p_n 的方程. 它们是

$$\begin{aligned} &m\lambda p_0 = \mu p_1, \\ &\{(m-n)\lambda + n\mu\} p_n = (m-n+1)\lambda p_{n-1} \\ &\qquad + (n+1)\mu p_{n+1} \quad (1 \leqslant n < r), \\ &\{(m-n)\lambda + r\mu\} p_n = (m-n+1)\lambda p_{n-1} \\ &\qquad + r\mu p_{n+1} \quad (r \leqslant n \leqslant m). \end{aligned} \tag{7.20}$$

从第一个方程我们得到比值 p_1/p_0. 由第二个方程用归纳法对于 $n < r$ 可得

$$(n+1)\mu p_{n+1} = (m-n)\lambda p_n; \tag{7.21}$$

最后，对于 $n \geqslant r$，由 (7.20) 中的最后一个方程可得

$$r\mu p_{n+1} = (m-n)\lambda p_n. \tag{7.22}$$

由这些方程可以依次算出比值 p_n/p_0. 最后，p_0 可由条件 $\Sigma p_k = 1$ 推出. 表 3 中的值就是用这种方法得到的.

我们比较表 2 与表 3 就可看出一些惊人的事实. 我们注意，这两个表所参照的是相同的机器 $(\lambda/\mu = 0.1)$，但在第二种情况，我们有 $m = 20$ 台机器及 $r = 3$ 个修理工. 每个修理工所照管的机器数已由 6 增加到 $6\frac{2}{3}$，而与此同时，机器更有效地得到维修，让我们用

$$\frac{w}{m} = \frac{\text{待修机器的平均数}}{\text{机器数}} \tag{7.23}$$

表 3

当 $\lambda/\mu=0.1$，$m=20$，$r=3$ 时的概率 p_n

n	被维修的机器数	待维修的机器数	闲着的修理工数	p_n
0	0	0	3	0.13625
1	1	0	2	0.27250
2	2	0	1	0.25888
3	3	0	0	0.15533
4	3	1	0	0.08802
5	3	2	0	0.04694
6	3	3	0	0.02347
7	3	4	0	0.01095
8	3	5	0	0.00475
9	3	6	0	0.00190
10	3	7	0	0.00070
11	3	8	0	0.00023
12	3	9	0	0.00007

来定义机器的损失系数，用

$$\frac{\rho}{r}=\frac{\text{闲着的修理工的平均数}}{\text{修理工数}} \tag{7.24}$$

来定义修理工的损失系数．为了实用的目的，我们可以把概率 $P_n(t)$ 及其极限 p_n 等同起来．在表 3 中我们有 $w=p_4+2p_5+3p_6+\cdots+17p_{20}$ 及 $\rho=3p_0+2p_1+p_2$. 表 4 明确地证明，对于我们的特定机器（对于它们 $\lambda/\mu=0.1$）来说，每二十台机器三个修理工要比每六台机器一个修理工更经济一些．第 449 页脚注提到的帕姆的表使我们能够求出每台机器的修理工的最经济的比率．

表 4

例 (c) 与 (d) 中所讨论的两种维修制的效率的比较

	I	II
机器数	6	20
修理工数	1	3
每个修理工的机器数	6	6⅔
修理工的损失系数	0.4845	0.4042
机器的损失系数	0.0549	0.01694

(e) **能量供应问题**[1). 设一条电路供应间歇地用电的 a 个焊工. 如果一个焊工在时刻 t 用电, 则他在时刻 $t+h$ 停止用电的概率为 $\mu h+o(h)$; 如果在时刻 t 他不用电, 则他在时刻 $t+h$ 以前需用电的概率为 $\lambda h+o(h)$. 各个焊工彼此独立地工作.

如果 n 个焊工在用电, 则我们说系统处于状态 E_n. 于是我们仅有有限个状态 $E_0, \cdots, E_a$.

如果系统处于状态 E_n, 则 $a-n$ 个焊工不用电, 且在时间 h 内增加一个新的用电者的概率为 $(a-n)\lambda h+o(h)$; 另一方面, n 个焊工之一停止用电的概率为 $n\mu h+o(h)$. 因此我们得到一个生灭过程, 其中

$$\lambda_n=(a-n)\lambda, \quad \mu_n=n\mu, \quad 0\leqslant n\leqslant a. \tag{7.25}$$

基本微分方程成为 $(1\leqslant n\leqslant a-1)$

$$\begin{aligned} P_0'(t) &= -a\lambda P_0(t)+\mu P_1(t), \\ P_n'(t) &= -\{n\mu+(a-n)\lambda\}P_n(t)+(n+1)\mu P_{n+1}(t) \\ &\quad +(a-n+1)\lambda P_{n-1}(t), \\ P_a'(t) &= -a\mu P_a(t)+\lambda P_{a-1}(t). \end{aligned} \tag{7.26}$$

容易验证, 极限概率由二项分布

$$p_n=\binom{a}{n}\left(\frac{\lambda}{\lambda+\mu}\right)^n\left(\frac{\mu}{\lambda+\mu}\right)^{a-n} \tag{7.27}$$

给出, 这个结果是可以根据直观的理由而预料到的.

8. 向后(瞻后)方程

在以上各节中我们研究了在时刻 t 系统处于状态 E_n 的概率 $P_n(t)$. 这个记号是方便的, 但由于它没有提到系统在时刻零的初始状态 E_i, 故易引起混淆. 所以为了理论上的目的引入记号 $P_{in}(t)$ 是更自然的; 这是已知系统在时刻零处于状态 E_i 的条件下, 它在时刻 t 处于状态 E_n 的概率. $P_{in}(t)$ 称为转移概率.

必须强调, 我们一直在研究这些转移概率, 且除记号外, 没有

1) 这个例子是根据 [66] 中所研究的问题(论述不够完善)而提出的.

任何改变．如果已知初始状态为 E_i，则 $\{P_{in}(t)\}$ 是在时刻 t 的绝对概率分布．如果在时刻零我们仅有初始状态的一个概率分布 $\{q_i\}$，则在时刻 t，E_n 的概率为

$$Q_n(t) = \sum_i q_i P_{in}(t). \tag{8.1}$$

在纯增殖过程与生灭过程的情况，我们曾看到，对于任意固定的 i 转移概率 $P_{in}(t)$ 满足基本微分方程 (3.2) 与 (5.2)．下标 i 仅在初始条件中出现，后者现在应当写为

$$P_{in}(0) = \begin{cases} 1, & \text{对于 } n = i, \\ 0, & \text{其它情况.} \end{cases} \tag{8.2}$$

这些基本微分方程是用将时间区间 $(0, t)$ 延长到 $(0, t+h)$ 并考虑在一段短时间中的改变的方法而导出的．我们也能将区间 $(0, t)$ 沿过去的方向延长并考虑 $(-h, 0)$ 中的改变．用这种方法我们得到一组新的微分方程，其中 n（代替 i）保持固定．

首先考虑纯增殖过程的情况，并且我们略去其概率以比 h 更快的速度趋向于 0 的事件．如果系统在时刻 $-h$ 处于 $E_i (i > 0)$ 而在时刻 t 转移到 E_n，则在时刻 0 它以概率 $1 - o(h)$ 处于 E_i 或 E_{i+1}．根据第 2 节与第 3 节中的方法，我们有

$$P_{in}(t+h) = P_{in}(t)(1 - \lambda_i h) + P_{i+1,n}(t)\lambda_i h + o(h). \tag{8.3}$$

因此对于 $i > 0$ 新的基本微分方程组具有如下形式

$$P'_{in}(t) = -\lambda_i P_{in}(t) + \lambda_i P_{i+1,n}(t), \tag{8.4}$$

及

$$P'_{0n}(t) = -\lambda_0 P_{0n}(t). \tag{8.5}$$

这些方程称为向后方程，为区别起见，我们称方程 (3.2) 为向前方程．初始条件为 (8.2)．

在生灭过程的情况，如果在时刻 $-h$ 系统处于状态 E_i，则在时刻零它应当处于 E_{i+1}，E_i 或 E_{i-1}，于是用同样的方法就可导出向后方程

$$P'_{in}(t) = -(\lambda_i + \mu_i)P_{i,n}(t) + \lambda_i P_{i+1,n}(t) + \mu_i P_{i-1,n}(t). \tag{8.6}$$

这些方程对应于 (5.2)．

显然，向前与向后方程不是相互独立的. 除在解不唯一的罕见情况外，向后方程的带有初始条件(8.2)的解自动满足向前方程. 这些联系此处仅作为下节中一般理论的准备而提及.

例. 普阿松过程. 在第2节中我们已将普阿松表达式(1.1)解释为在任意长为 t 的时间间隔中恰好来到 n 个呼唤的概率. 如果在从0到 t 的时间间隔中恰好来到 n 个呼唤，则我们说在时刻 t 系统处于状态 E_n. 由在时刻 t_1 的 E_i 转移到在时刻 t_2 的 E_n 意味着在 (t_1, t_2) 中来到 $n-i$ 个呼唤. 这仅当 $n \geq i$ 时才有可能，因此对于普阿松过程的转移概率我们有

$$\begin{aligned} P_{in}(t) &= e^{-\lambda t}\frac{(\lambda t)^{n-i}}{(n-i)!}, \quad \text{当 } n \geq i, \\ P_{in}(0) &= 0, \qquad\qquad\qquad \text{当 } n < i. \end{aligned} \tag{8.7}$$

向前与向后方程分别为

$$P'_{in}(t) = -\lambda P_{in}(t) + \lambda P_{i,n-1}(t) \tag{8.8}$$

及

$$P'_{in}(t) = -\lambda P_{in}(t) + \lambda P_{i+1,n}(t), \tag{8.9}$$

容易验证，(8.7)是这两个方程组的解，且满足初始条件(8.2).

9. 推广；柯尔莫果洛夫方程

到目前为止，我们仅讨论了从状态 E_n 仅能直接转移到其邻近状态 E_{n+1} 与 E_{n-1} 的过程. 而且，过程在时间上是均匀的，即转移概率对于所有长为 t 的时间间隔是相同的. 现在我们考虑更一般的过程，其中这两个假设都被放弃.

象在通常马尔科夫链理论中那样，我们将允许从任意状态 E_i 直接转移到任意状态 E_n. 转移概率可以随时间变化. 这时仅指定时间间隔的长度是不够的，而必须指定它的两个端点. 因此，我们用 $P_{in}(\tau, t)$ 来表示已知系统在时刻 τ 的状态为 E_i 的条件下，它在其后的时刻 t 处于状态 E_n 的条件概率. 仅当 $\tau < t$ 时记号 $P_{in}(\tau, t)$ 才有意义. 如果过程在时间上是齐次的，则 $P_{in}(\tau, t)$ 仅

依赖于差 $t-\tau$，于是我们可用 $P_{in}(t)$（它不依赖于 τ）来代替 $P_{in}(\tau, \tau+t)$.

我们的过程的主要性质是第十五章第 10 节中所讨论过的马尔科夫性：在任意一个时刻系统的状态已给定的情况下，系统将来的变化与过去无关. 更精确地说，考虑三个时刻 $\tau < s < t$，并假定在时刻 τ 系统处于状态 E_i 且在时刻 s 处于状态 E_ν. 对于一个任意的过程来说，系统在时刻 t 处于状态 E_n 的(条件)概率依赖于 i 与 ν；换句话说，不仅"现在的状态" E_ν 对在时刻 t 的状态有影响，而且过去的状态对它也有影响. 然而，对于马尔科夫过程来说情况就不是这样. 对于这种过程，所考虑的概率等于 $P_{\nu n}(s, t)$，即从时刻 s 的 E_ν 转移到时刻 t 的 E_n 的概率. 在时刻 $\tau < s$ 系统处于状态 E_i 的知识对将来不产生影响. 由这个假设可以直接导出重要的结论. 由时刻 τ 的 E_i 转移到时刻 t 的 E_n 必须经过时刻 s 的某个状态，对于马尔科夫过程来说，经过一个特定状态 E_ν 的概率为 $P_{i\nu}(\tau, s)P_{\nu n}(s, t)$. 由此我们必须有

$$P_{in}(\tau, t) = \sum_\nu P_{i\nu}(\tau, s)\, P_{\nu n}(s, t) \tag{9.1}$$

对一切 $\tau < s < t$ 成立. 这是查普曼-柯尔莫果洛夫方程. 这个方程与第十五章的方程(10.3)相当；前者是连续时间参数的情况，后者仅当时间参数取整数值时成立.

在第十五章第 10 节中我们曾证明，查普曼-柯尔莫果洛夫方程并不对所有的随机过程成立. 为了我们的目的，我们可用(9.1)来规定我们所涉及的过程类[1]. 事实上，我们将仅增加正规限制并由(9.1)导出我们的基本微分方程. 查普曼-柯尔莫果洛夫方程的导出有其概率背景，但我们不需参考它；一旦给出了(9.1)，我们就能容易地确定概率 $P_{in}(t)$ 的微分方程并用纯分析的方法进行讨论.

---

1) 柯尔莫果洛夫方程是否刻划了马尔科夫过程的问题，提出了需要研究现实样本函数 $\mathbf{X}(t)$ 的困难问题. 应当记住，我们是在用一种简捷的方法来得出关于某些概率的微分方程，而并未从各个方面对过程进行分析.

在时间均匀的过程的情况，方程(9.1)具有如下较简单的形式

$$P_{in}(t+s)=\sum_{\nu}P_{i\nu}(t)P_{\nu n}(s). \tag{9.2}$$

对于普阿松过程这个方程化归为普阿松分布的褶积性质 [第十一章例 (2.c)].

现在我们引入我们的基本正规条件，它显然是生灭过程的初始假设的推广.

假设 1. 对于每个状态 E_n 有一个连续函数 $c_n(t)\geqslant 0$ 与之对应，使得当 $h\to 0$ 时有

$$\frac{1-P_{nn}(t,t+h)}{h}\to c_n(t). \tag{9.3}$$

(9.3) 的概率解释是显然的；如果在时刻 t 系统处于状态 E_n，则在 $(t,t+h)$ 中出现一个改变的概率为 $c_n(t)h+o(h)$. 分析上，关系 (9.3) 要求，当 $s\to t$ 时 $P_{nn}(t,s)\to 1$ 以及 $P_{nn}(t,x)$ 在 $x=t$ 有导数. 函数 $c_n(t)$ 起着生灭过程中 $\lambda_n+\mu_n$ 的作用. 在时间均匀的过程的情况，c_n 不依赖于 t.

假设 2. 对于每对状态 E_j, $E_k(j\neq k)$ 有转移概率 $p_{jk}(t)$（依赖于时间）与之对应，使得当 $h\to 0$ 时有

$$\frac{P_{jk}(t,t+h)}{h}\to c_j(t)p_{jk}(t)\quad (j\neq k). \tag{9.4}$$

$p_{jk}(t)$ 对于 t 连续，且对于每个固定的 t, j

$$\sum_k p_{jk}(t)=1,\ p_{jj}(t)=0. \tag{9.5}$$

这里 $p_{jk}(t)$ 可解释为在 $(t,t+h)$ 中一个改变自 E_j 出现的条件下，这个改变使系统由 E_j 转移到 E_k 的条件概率. 在生灭过程中

$$p_{j,j+1}(t)=\frac{\lambda_j}{\lambda_j+\mu_j},\quad p_{j,j-1}(t)=\frac{\mu_j}{\lambda_j+\mu_j}, \tag{9.6}$$

对于 j 与 k 的所有其它组合 $p_{jk}(t)=0$. 对于每个固定的 t, $p_{jk}(t)$ 可以解释为马尔科夫链的转移概率.

这两个假设足够导出 $P_{jk}(\tau, t)$ 的向后方程组，但对于向前方程我们需要补充

假设 3. 对于固定的 k，(9.4) 中的极限关于 j 一致地成立.

这个假设的必要性对于无穷微分方程组的理论有重要意义，我们将在下一节讨论.

下面我们来推导 $P_{ik}(\tau, t)$（作为 t 与 τ 的函数）的微分方程. 由方程 (9.1) 我们有

$$P_{ik}(\tau, \tau+h) = \sum_j P_{ij}(\tau, t) P_{jk}(t, t+h). \tag{9.7}$$

根据 (9.3) 来表示右边的项 $P_{kk}(t, t+h)$，我们得

$$\begin{aligned}\frac{P_{ik}(\tau, t+h) - P_{ik}(\tau, t)}{h} &= -c_k(t)P_{ik}(\tau, t) \\ &\quad + \frac{1}{h}\sum_{j\neq k} P_{ij}(\tau, t)P_{jk}(t, t+h) + \cdots,\end{aligned} \tag{9.8}$$

其中略去的项随 h 趋向于 0，和是对除 $j = k$ 外的所有 j 来求的. 现在我们能将 (9.4) 应用于和中的项. 因为(由假设 3) 趋向于极限对所有的 j 是一致的，故右边有一极限. 所以左边也有极限，这意味着 $P_{ik}(\tau, t)$ 具有对于 t 的偏导数，且

$$\frac{\partial P_{ik}(\tau, t)}{\partial t} = -c_k(t)P_{ik}(\tau, t) + \sum_j P_{ij}(\tau, t)c_j(t)p_{jk}(t). \tag{9.9}$$

这是基本的向前微分方程组. 注意，i 与 τ 是固定的，故我们有(尽管形式上出现偏导数)关于无限多个函数 $P_{ik}(\tau, t)$ ($k = 0, 1, 2, \cdots$) 的常微分方程组. 参数 i 与 τ 仅在初始条件中出现

$$P_{ik}(\tau, \tau) = \begin{cases}1, & \text{当 } k = i, \\ 0, & \text{其它情况.}\end{cases} \tag{9.10}$$

向后微分方程组可用类似的方法得到，而且推导实际上更为简单，因为我们完全不需要假设 3. 至于方程 (9.3) 与 (9.4)，采用形式

$$\frac{1 - P_{nn}(t-h, t)}{h} \to c_n(t) \tag{9.3a}$$

$$\frac{P_{jk}(t-h,t)}{h} \to c_j(t)p_{jk}(t) \quad (j \neq k). \tag{9.4a}$$

能够证明这些关系与(9.3)及(9.4)等价，但我们将简单地以(9.3a)，(9.4a)与(9.5)作为我们的基本假设而着手讨论. 将查普曼-柯尔莫果洛夫方程改写为如下形式

$$P_{ik}(\tau-h,\, t) = \sum_{\nu} P_{i\nu}(\tau-h,\, \tau)P_{\nu k}(\tau,\, t), \tag{9.11}$$

利用(9.3a)(其中 $n=i$)，我们得到

$$\frac{P_{ik}(\tau-h,\, t)-P_{ik}(\tau,\, t)}{h} = -c_i(\tau)P_{ik}(\tau,\, t)$$

$$+\frac{1}{h}\sum_{\nu\neq i} P_{i\nu}(\tau-h,\, \tau)P_{\nu k}(\tau,\, t)+\frac{o(h)}{h}. \tag{9.12}$$

这里 $h^{-1}P_{i\nu}(\tau-h,\, \tau) \to c_i(\tau)\, p_{i\nu}(\tau)$，且(9.12)右边的和中的极限总是一致收敛的. 事实上，如果 $N>i$，则我们有

$$0 \leqslant h^{-1}\sum_{\nu=N+1}^{\infty} P_{i\nu}(\tau-h,\, \tau)P_{\nu k}(\tau,\, t)$$

$$\leqslant h^{-1}\sum_{\nu=N+1}^{\infty} P_{i\nu}(\tau-h,\, \tau) \leqslant h^{-1}\left\{1-\sum_{\nu=0}^{N} P_{i\nu}(\tau-h,\, \tau)\right\}$$

$$\to c_i(\tau)\left\{1-\sum_{\nu=0}^{N} p_{i\nu}(\tau)\right\}. \tag{9.13}$$

由条件(9.5)，选取充分大的 N 就可使上式右边任意小. 由此可知，在(9.12)中逐项取极限是容许的，于是我们得到

$$\frac{\partial P_{ik}(\tau,\, t)}{\partial \tau} = c_i(\tau)P_{ik}(\tau,\, t) - c_i(\tau)\sum_{\nu} p_{i\nu}(\tau)P_{\nu k}(\tau,\, t). \tag{9.14}$$

上式，连同初始条件(9.10)，是基本的向后微分方程组.

这两组微分方程是由柯尔莫果洛夫(他奠定了马尔科夫过程理论的基础)首先导出的[67]. 已经证明[1)]，这两组方程总存在一个公共解 $\{P_{ik}(\tau,\, t)\}$，它满足查普曼-柯尔莫果洛夫方程(9.1)且

---

1) 见第442页脚注1.

$$P_{ik}(\tau, t) \geqslant 0, \quad \sum_k P_{ik}(\tau, t) \leqslant 1. \tag{9.15}$$

由纯增殖过程（第 4 节）可知，$P_{ik}(\tau, t)$ 之和不必为 1，差 $1 - \sum P_{ik}(\tau, t)$ 说明在有限的时间间隔 (τ, t) 中出现无限多个转移的可能性．如果 $\sum P_{ik}(\tau, t) = 1$，则解 $\{P_{ik}(\tau, t)\}$ 是唯一的，但一般地说，不同的过程可以满足同一向前与向后方程（见第 10 节）．从应用的观点来说，不等式 $\sum P_{ik}(\tau, t) < 1$ 的可能性可以放心地不予考虑．

例．广义普阿松过程．考虑所有 $c_i(t)$ 等于同一常数 λ 且 p_{jk} 不依赖于 t 的情况．在这种情况中 p_{jk} 是一个通常马尔可夫链的转移概率，并且（如第十五章中那样）我们用 $p_{jk}^{(n)}$ 表示它的高阶转移概率．

由 $c_i(t) = \lambda$ 可以推出，在间隔 $(t, t+h)$ 中出现一个转移的概率不依赖于系统在时刻 t 的状态且等于 $\lambda h + o(h)$．这蕴含在间隔 (τ, t) 中的转移数具有参数为 $\lambda(t-\tau)$ 的普阿松分布．在已知恰好有 n 次转移出现的条件下，由 j 转移到 k 的（条件）概率为 $p_{jk}^{(n)}$．因此

$$P_{ik}(\tau, t) = e^{-\lambda(t-\tau)} \sum_{n=0}^{\infty} \frac{\lambda^n (t-\tau)^n}{n!} p_{ik}^{(n)} \tag{9.16}$$

（这里，和通常一样，$p_{jj}^{(0)} = 1$，当 $j \neq k$ 时，$p_{jk}^{(0)} = 0$）．容易验证，(9.16) 事实上是两个微分方程组 (9.9) 与 (9.14) 满足边界条件 (9.10) 的一组解．

特别地，如果

$$\text{当 } k < j \text{ 时 } p_{jk} = 0, \quad \text{当 } k \geqslant j \text{ 时 } p_{jk} = f_{k-j}, \tag{9.17}$$

则 (9.16) 化归为第十二章第 1 节中的复合普阿松分布．

10. 病 态 过 程

纯增殖过程的例子（第 3，4 节）证明，由柯尔莫果洛夫微分方程所确定的转移概率 $P_{ik}(t)$ 之和并不必须为 1；如下情况是可能

发生的:

$$\sum_k P_{ik}(t) < 1. \tag{10.1}$$

这种现象是 1940 年发现的,当时它是出人意料之外并使人感到麻烦的. 大量的文献对它以及柯尔莫果洛夫微分方程并不总确定唯一的一组转移概率 $P_{ik}(t)$ 这一有关事实进行了研究. 具有这些性质的过程通常称为病态的[1]. 后来认识到,此处的情况实际上是扩散理论中常见情况的一个简单而自然的类似, 将有助于更好地理解这种现象. (10.1) 的出现不再使人感到麻烦, 而是导致了柯尔莫果洛夫微分方程的理论与扩散理论具有相同的基本特征这一事实的发现. 尽管基本方程及所用的分析工具的外貌完全不同, 但在两种理论中我们遇到相同类型的边界条件及其它一些类似; 每种理论都可以借助另一个而更好地得到理解, 因此在两者之间实际上不能划出鲜明的界线. 在这方面马尔科夫过程理论已达到意想不到的和令人满意的内部统一.

让我们重新考虑第 3 节中的简单增殖过程. 系统在初始状态 E_0 花费一些时间,然后从那里移动到 E_1,在那里停一会儿,再移动到 E_2, 等等. 在 E_0 逗留的时间超过 t 的概率 $P_0(t)$ 可由 (3.2) 求出: $P_0(t) = e^{-\lambda_0 t}$. 这个逗留时间 $\mathbf{T}_0$ 是一随机变量, 但它的值域是正 t 轴,所以超出了本书的范围. 不过,由于从几何分布过渡到指数分布是平凡的,因此我们可以稍微越出这个范围而不必担心. 用具有几何分布的离散随机变量来逼近 $\mathbf{T}_0$ 表明,用

$$\mathbf{E}(\mathbf{T}_0) = \int_0^\infty te^{-\lambda_0 t}dt = \lambda_0^{-1}. \tag{10.2}$$

来定义在 E_0 的期望逗留时间是自然的. 在系统进入 E_j 的时刻, 状态 E_j 起着初始状态的作用, 于是同样的论断适用于在 E_j 的逗留时间 $\mathbf{T}_j$: 在 E_j 的期望逗留时间 $\mathbf{E}(\mathbf{T}_j) = \lambda_j^{-1}$. 由此可知, $\lambda_0^{-1} + \lambda_1^{-1} + \cdots + \lambda_n^{-1}$ 是系统经过 $E_0, E_1, \cdots, E_n$ 所需的时间的期望值,我们可将第 4 节中的准则重述如下:

---

1) 在初版中本节的标题为"退化过程".

$\sum P_n(t)=1$ 对一切 t 成立的充要条件是

$$\sum \mathbf{E}(\mathbf{T}_j)=\sum \lambda_j^{-1}=\infty; \tag{10.3}$$

即在 E_0, E_1, E_2, $\cdots$ 所花费的总时间的期望值必须为无限．当然，$L_0(t)=1-\sum P_n(t)$ 是系统在时刻 t 以前已经过所有的状态的概率．

在这种形式中定理表面上是很合理的． 如果在 E_j 的期望逗留时间为 2^{-j}，则系统在时间 $1+2^{-1}+2^{-2}+\cdots=2$ 中经过所有的状态的概率必须为正．类似，沿 x 轴运动且速度按指数增加的质点在有限的时间内横过整个轴．

如果用增殖过程作为总体增长的模型，则状态 E_n 代表总体的实际大小 n，而在有限的时间内达到无穷则表示一种爆炸．在这时 (10.1) 确实代表一种异常现象，但对于其它的应用它却可以作为一种正常状态而出现．从几何上说，没有理由把状态 E_0, E_1, E_2, $\cdots$ 置于 x 轴上的点 $0, 1, 2, \cdots$ 处．设想 E_n 被置于 x 轴上的点 x_n 处，其中 $0=x_0<x_1<x_2<\cdots<\cdots$ 且 $x_n\to 1$．于是增殖过程可用“质点”的运动来描述，这个质点从 $x_0=0$ 出发，然后跳到 x_1，一会儿后又跳到 x_2，等等．在这种描述中，经过所有状态的质点已到达极限点 1；这个事件在有限的时间中出现是自然的．将概率运动与从原点出发、在位置 x 处具有速度 $f(x)$ 的质点的决定性运动相比较．它在时刻 t 的位置 $x(t)$ 满足微分方程 $x'(t)=f(x(t))$ 且到达点 1 的时间 τ 为

$$\tau=\int_0^1 \frac{dx}{f(x)}\leqslant\infty. \tag{10.4}$$

点 1 实际上能否在有限的时间内到达（或仅仅渐近地趋近）依赖于速度倒数的积分的敛散性． 在概率模型中运动是跳跃式进行的，λ_n^{-1} 为由 x_n 到 x_{n+1} 所需的平均时间． 从这个观点来看，(10.3) 与 (10.4) 象是一对孪生准则．

让我们继续考虑简单的增殖过程，并说明准则 (10.3) 怎样关系到柯尔莫果洛夫微分方程的唯一性问题．

作为转移概率来考虑．第 3 节中的 $P_n(t)$ 应当写为 $P_{in}(t)$．基

本微分方程(3.2)同样适用于 $P_{ik}(t)$，其中 i 是任意的(但是固定的)；于是我们有向前微分方程

$$P'_{i0}(t) = -\lambda_0 P_{i0}(t),$$
$$P'_{ik}(t) = -\lambda_k P_{ik}(t) + \lambda_{k-1} P_{i,k-1}(t), \tag{10.5}$$

其中 $i \geqslant 0$ 是固定的而 $k = 1, 2, \cdots$. 在(8.4)与(8.5)中我们有对应的向后方程

$$P'_{ik}(t) = -\lambda_i P_{ik}(t) + \lambda_i P_{i+1,k}(t), \tag{10.6}$$

其中 $k \geqslant 0$ 是固定的，$i = 0, 1, 2, \cdots$. 初始条件显然为

$$P_{ii}(0) = 1,\ P_{ik}(0) = 0,\ k \neq i. \tag{10.7}$$

由(10.5)易知，$P_{i0}(t)$ 被第一个微分方程及边界条件(10.7)唯一地确定. 于是对于 $k = 1, 2, \cdots$，我们能依次计算 $P_{ik}(t)$. 如何解一阶线性方程是熟知的，对于向前方程(10.5)满足初始条件(10.7)的唯一解，我们有如下容易验证的公式：

$$P_{ik}(t) = 0\ \ (k < i),\ P_{ii}(t) = e^{-\lambda_i t}$$
$$P_{ik}(t) = \lambda_{k-1}\int_0^t e^{-\lambda_k s} P_{i,k-1}(t-s)ds,\ \ (k > i). \tag{10.8}$$

对于向后方程(10.6)情况完全不同. 我们将证明，"当 $\sum \lambda_n^{-1} < \infty$ 时解的唯一性不成立.

引理. 由(10.8)给出的向前方程的唯一解 $P_{ik}(t)$ 自动满足向后方程(10.6). 如果 $\bar{P}_{ik}(t)$ 是(10.6)—(10.7)的任一非负解，则

$$\bar{P}_{ik}(t) \geqslant P_{ik}(t). \tag{10.9}$$

证. 在(10.6)中，在所有的 $P_{ik}(t)$ 上面加上一条短线. 第 i 个方程可以作为 $\bar{P}_{ik}(t)$ 的一个线性微分方程来解，我们得到

$$\bar{P}_{ik}(t) = \lambda_i \int_0^t e^{-\lambda_i s}\bar{P}_{i+1,k}(t-s)ds,\ k \neq i$$
$$\bar{P}_{kk}(t) = e^{-\lambda_k t} + \lambda_k \int_0^t e^{-\lambda_k s}\cdot \bar{P}_{k+1,k}(t-s)ds. \tag{10.10}$$

[注意：这不是一组递推公式，因此不能用来解方程组(10.6).]

设 $P_{ik}(t)$ 代表由(10.8)给出的向前方程的解. 对于每个 k 及 $i > k$，令 $\bar{P}_{ik}(t) = P_{ik}(t) = 0$. 这些函数满足(10.10). 而且由(10.10)有 $\bar{P}_{kk}(t) = e^{-\lambda_k t} = P_{kk}(t)$. 在(10.10)中依次令 $i =$

$k-1, k-2, \cdots$，我们就对所有的 i 确定了 $\bar{P}_{ik}(t)$，且它们是向后方程(10.6)满足初始条件(10.7)的一个解．显然 $\bar{P}_{k-1,k}(t)=P_{k-1,k}(t)$．我们将用归纳法验证 $\bar{P}_{ik}(t)=P_{ik}(t)$．设当 $k-i\leqslant r$（此处 $r\geqslant 1$ 是一整数）时这个等式成立（我们知道当 $r=1$ 时这个假设是成立的），并令 $k-i=r+1$．(10.10) 中的积分用 $\bar{P}_{i+1,k}=P_{i+1,k}$ 来表示 $\bar{P}_{ik}$，而(10.8)又将 $P_{i+1,k}$ 表示为包含 $P_{i+1,k-1}=\bar{P}_{i+1,k-1}$ 的积分．于是 $\bar{P}_{ik}$ 可用 $\bar{P}_{i+1,k-1}$ 的二重积分来表示．颠倒积分次序我们得

$$\begin{aligned}\bar{P}_{ik}(t) &= \lambda_i\,\lambda_{k-1}\int_0^t e^{-\lambda_k x}dx\int_0^{t-x}e^{-\lambda_i s}\bar{P}_{i+1,k-1}(t-x-s)ds\\&=\lambda_{k-1}\int_0^t e^{-\lambda_k x}\bar{P}_{i,k-1}(t-x)dx. \qquad (10.11)\end{aligned}$$

根据归纳法假设 $\bar{P}_{i,k-1}(t)=P_{i,k-1}(t)$，将(10.8)与(10.11)相比较即可证明 $\bar{P}_{ik}(t)=P_{ik}(t)$．

剩下所要证明的是，(10.9) 对于向后方程 (10.6) 的任意解 $\bar{P}_{ik}(t)\geqslant 0$ 成立．现在 $\bar{P}_{ik}(t)$ 与 $P_{ik}(t)$ 都满足 (10.10)．对于 $i>k$ 我们有 $\bar{P}_{ik}(t)\geqslant P_{ik}(t)=0$．依次令 $i=k, k-1, k-2, \cdots$ 即可证明(10.9)对所有的 i, k 成立，于是引理得证．

分以下两种情况来讨论．

(a) $\sum\lambda_n^{-1}=\infty$ 的情况．由第 4 节可知，在这种情况 $\sum\limits_k P_{ik}(t)=1$．由此可知，向后方程的任何其它正解之和必大于 1，而这是不容许的．因此在这种情况向前与向后方程的容许解都是唯一的．这个公共解表示增殖过程满足条件 $\sum P_{ik}(t)=1$ 的转移概率．（容易用微分法证明查普曼-柯尔莫果洛夫方程(9.1)成立．）

(b) $\sum\lambda_n^{-1}<\infty$ 的情况．我们知道，在这种情况有 $\sum P_{ik}(t)<1$．于是

$$L_i(t)=1-\sum_{k=0}^{\infty}P_{ik}(t) \qquad (10.12)$$

为从 E_i 出发、"无限性"在时刻 t 之前达到的概率．我们知道，

$\bar{P}_{ik}(t) = P_{ik}(t)$ 满足 (10.10), 又通过求和我们看到

$$L_i(t) = \lambda_i \int_0^t e^{-\lambda_i s} L_{i+1}(t-s) ds \tag{10.13}$$

或

$$L_i'(t) = -\lambda_i L_i(t) + \lambda_i L_{i+1}(t) \quad L_i(0) = 0. \tag{10.14}$$

在这种情况可以推出, 无穷微分方程组 (10.14) 具有满足条件 $L_i(0) = 0$ 的非零解. 对于任意的 $A_k(t)$ 矩阵

$$\bar{P}_{ik}(t) = P_{ik}(t) + L_i(t)A_k(t) \tag{10.15}$$

是向后方程 (10.6) 满足初始条件 (10.7) 的一个解.

现在的问题是能否确定 $A_k(t)$, 使得 $\bar{P}_{ik}(t)$ 成为满足查普曼-柯尔莫果洛夫方程 (9.1) 的转移概率, 答案是肯定的. 此处我们不准备证明这个论断但将给出一个概率解释.

$P_{ik}(t)$ 定义所谓吸收壁过程: 当系统达到无穷时, 过程终止. 杜勃[68]最先研究了一种返回过程, 在其中, 一旦到达无穷, 系统立即回到 E_0 (或其它某个指定状态)且过程从头开始. 在这种过程中系统由 E_0 到 E_5 可以经过五步或无穷多步. 这个过程的转移概率具有 (10.5) 的形式. 它们满足向后方程 (10.6), 但不满足向前方程 (10.5).

这说明为什么在向前方程的推导中我们不得不引入显得有些奇怪的假设了, 它对于向后方程并不是必需的: 在概率和直观上都简单的假设 1—2 是与返回过程相一致, 对于这种过程向前方程不成立. 换句话说, 如果我们从假设 1—2 出发, 则柯尔莫果洛夫向后方程成立, 但对于向前方程则必须加上另一项[1].

纯增殖过程太平凡, 没有真正的趣味, 但这些条件对于柯尔莫果洛夫方程的最一般的情况是典型的. 不过, 有两个本质上的新现象出现. 首先, 增殖过程仅含有一条逸向"无穷"的路线, 或用抽象的术语来说, 一个单一的边界点. 与此相对照, 一般过程可能包含具有复杂拓扑结构的边界. 第二, 在增殖过程中, 由于仅转移

---

1) [55] 中给出了它的形式, 在这篇论文中研究了各种类型的过程及适当的边界条件.

$E_n \to E_{n+1}$ 为可能，故运动是指向边界的. 能构造一个不同类型的过程；例如，可将方向颠倒而得到一个过程，在其中仅转移 $E_{n+1} \to E_n$ 是可能的. 这种过程不是在边界结束而是从那里发源. 在生灭过程中,沿两个方向的转移都是可能的,正如在一维扩散中那样. 在这种情况，存在与扩散理论中的弹性与反射壁过程相类似的过程,但它们的描述超出本书的范围.

11. 问　　题

1. 在由 (3.2) 定义的纯增殖过程中,对于所有的 n 令 $\lambda_n > 0$. 证明对于每个固定的 n 函数 $P_n(t)$ 首先增加然后递减到 0. 如果 t_n 是极大值的位置，则 $t_1 < t_2 < t_3 < \cdots$. 提示：利用归纳法;微分 (3.2).

2. (续上). 如果 $\Sigma\lambda_n^{-1} = \infty$, 试证明 $t_n \to \infty$. 提示：如果 $t_n \to \tau$, 则对于固定的 $t > \tau$ 序列 $\lambda_n P_n(t)$ 增加. 利用 (4.10).

3. **尤尔过程.** 导出由 (3.4) 所定义的分布的均值与方差. [仅利用微分方程,而不用显式 (3.5).]

4. **纯消亡过程.** 试求仅有由 E_n 到 E_{n-1} 的转移的尤尔型过程的微分方程. 试求分布 $P_n(t)$ 及其均值与方差,假定初始状态为 i.

5. **停车场.** 在一个有 N 个位置的停车场中来到的流是强度为 λ 的普阿松型流,但仅当有空位置可以采用时是这样. 试求恰有 n 个位置被占用的概率 $P_n(t)$ 所满足微分方程.

6. 在排队中最后来到的顾客首先受到服务[69]. 试求恰有 n 个新来者在随机选出的一个顾客的等待时间中受到服务的概率 $P_n(t)$ 所满足的微分方程.

7. **波里亚过程[70].** 这是一个非平稳的纯增殖过程，其参数依赖于时间:

$$\lambda_n(t) = \frac{1 + an}{1 + at}. \qquad (11.1)$$

试证明满足初始条件 $P_0(0) = 1$ 的解为

$$P_0(t) = (1 + at)^{-1/a},$$

$$P_n(t) = t^n (1 + at)^{-n-1/a} \frac{(1 + a)(1 + 2a)\cdots\{1 + (n - 1)a\}}{n!}. \qquad (11.2)$$

由微分方程证明均值与方差分别为 t 与 $t(1 + at)$.

8．（续上）． 波里亚过程能由第五章例（2.c）中的罐子模型取极限而得到．如果系统的状态定义为红球的数目，则在第 $(n+1)$ 次抽取时的转移概率为

$$p_{k,n}=\frac{r+kc}{r+b+nc}=\frac{p+k\gamma}{1+n\gamma}, \tag{11.3}$$

其中 $p=r/(r+b)$，$\gamma=c/(r+b)$。

象由伯努利试验过渡到普阿松分布一样，设抽球以每 h 个时间单位一次的速度进行，并令 $h\to 0$，$n\to\infty$ 使得 $np\to t$，$n\gamma\to at$。试证明在取极限时（11.3）导致（11.1）。并证明第五章的波里亚分布（2.3）化为（11.2）。

9．**线性增长．** 如果在由（5.7）定义的过程中 $\lambda=\mu$，$P_1(0)=1$，则

$$P_0(t)=\frac{\lambda t}{1+\lambda t},\quad P_n(t)=\frac{(\lambda t)^{n-1}}{(1+\lambda t)^{n+1}}. \tag{11.4}$$

最终灭绝的概率为 1。

10．（续上）． 取 $P_n(t)=A(t)\,B^n(t)$ 作为（5.7）的一个试探解，证明具有 $P_1(0)=1$ 的解为

$$\begin{aligned}&P_0(t)=\mu B(t),\\&P_n(t)=\{1-\lambda B(t)\}\{1-\mu B(t)\}\{\lambda B(t)\}^{n-1},\end{aligned} \tag{11.5}$$

其中

$$B(t)=\frac{1-e^{(\lambda-\mu)t}}{\mu-\lambda e^{(\lambda-\mu)t}}. \tag{11.6}$$

11．（续上）． 证明母函数 $P(s,t)=\Sigma P_n(t)s^n$ 满足偏微分方程

$$\frac{\partial P}{\partial t}=\{\mu-(\lambda+\mu)s+\lambda s^2\}\frac{\partial P}{\partial s}. \tag{11.7}$$

12．（续上）． 设 $M_2(t)=\Sigma n^2P_n(t)$ 及 $M(t)=\Sigma nP_n(t)$（如第 5 节中那样）。证明

$$M_2'(t)=2(\lambda-\mu)M_2(t)+(\lambda+\mu)M(t). \tag{11.8}$$

证明当 $\lambda>\mu$ 时 $\{P_n\}$ 的方差由

$$e^{2(\lambda-\mu)t}\{1-e^{(\mu-\lambda)t}\}(\lambda+\mu)/(\lambda-\mu) \tag{11.9}$$

给出。

13．对于过程（7.2）母函数 $P(s,t)=\Sigma P_n(t)s^n$ 满足偏微分方程

$$\frac{\partial P}{\partial t}=(1-s)\left\{-\lambda P+\mu\frac{\partial P}{\partial s}\right\}. \tag{11.10}$$

其解为

$$P = e^{-\lambda(1-s)(1-e^{-\mu t})/\mu}\{1-(1-s)e^{-\mu t}\}^i.$$

对于 $i=0$ 这是参数为 $\lambda(1-e^{-\mu t})/\mu$ 的普阿松分布。当 $t\to\infty$ 时，分布 $\{P_n(t)\}$ 趋向于参数为 λ/μ 的普阿松分布。

14. 对于由 (7.26) 定义的过程母函数 $P(s,t)=\Sigma P_n(t)s^n$ 满足偏微分方程

$$(\mu+\lambda s)\frac{\partial P}{\partial s}=a\lambda P, \tag{11.11}$$

它的一个解为 $P=\{(\mu+\lambda s)/(\lambda+\mu)\}^a$。

15. 在"最简单的占线问题"例 7(a) 中，令 $Q_n(t)$ 为系统从 E_n 出发，将在时刻 t 以前到达 E_0 的概率。证明 $Q_n(t)$ 满足微分方程

$$Q_n'(t)=-(\lambda+n\mu)Q_n(t)+\lambda Q_{n+1}(t)+n\mu Q_{n-1}(t),\quad (n\geqslant 2),$$

$$Q_1'(t)=-(\lambda+\mu)Q_1(t)+\lambda Q_2(t)+\mu \tag{11.12}$$

及初始条件 $Q_n(0)=0$。

16.（续上）。 对于由任意的一组向前方程所定义的过程考虑同一问题。证明 $Q_n(t)$ 满足对应的向后方程（对于固定的 k），其中 $P_{0k}(t)$ 用 1 代替。

17. 证明纯增殖过程的转移概率及生灭过程的转移概率都满足查普曼-柯尔莫果洛夫方程 (9.1)。

18. 设 $P_{ik}(t)$ 满足查普曼-柯尔莫果洛夫方程 (9.1)。设 $P_{ik}(t)>0$ 及 $S_i(t)=\sum_k P_{ik}(t)\leqslant 1$，证明或者对于所有的 t 有 $S_i(t)=1$ 或者对于所有的 t 有 $S_i(t)<1$。

19. **遍历性质。** 考虑具有有限多个状态的稳定过程；即假定微分方程组(9.9) 是有限的且 c_j 与 p_{jk} 为常数。证明解为指数项 $e^{\lambda(t-\tau)}$ 的线性组合，其中 λ 的实部是负的，除非 $\lambda=0$。证明转移概率的渐近性态和有限马尔科夫链中的情况一样，只是周期性情况是不存在的。

问 题 解 答

第 十 一 章

1. $sP(s)$ 与 $P(s^2)$。

2. (a) $(1-s)^{-1}P(s)$; (b) $(1-s)^{-1}sP(s)$; (c) $\{1-sP(s)\}/(1-s)$; (d) $p_0s^{-1}+\{1-s^{-1}P(s)\}/(1-s)$; (e) $\frac{1}{2}\{P(s^{1/2})+P(-s^{1/2})\}$.

3. $U(s)=pqs^2/(1-ps)(1-qs)$。均值 $=1/pq$，方差 $=(1-3pq)/p^2q^2$。

6. 一个零是第一、第二、第三、…个零，故 $U(s)=\Sigma F^k(s)$。

7. 母函数为 $\{1-F(s)\}(1-s)^{-1}=(1+s)U(s)$。

8. 母函数为 $\Sigma\left\{\frac{1}{2}F(s)\right\}^k=2F(s)s^{-2}-1$。

9. 相同的母函数。

10. 第 k 个零必须在第 $2r\leqslant n$ 次试验中出现，而随后 $n-2r$ 次试验不可产生一个零。

11. 对于 $P(1)<1$ 的情况利用显然类似于 (1.6) 的一个公式。

12. 利用 $\mathbf{X}_\nu$ 的几何分布的母函数，不需计算，我们有

$$P_r(s)=s^r\left(\frac{N-1}{N-s}\right)\left(\frac{N-2}{N-2s}\right)\cdots\left(\frac{N-r+1}{N-(r-1)s}\right).$$

13. $P_r(s)\{N-(r-1)s\}=P_{r-1}(s)(N-r-1)s$.

14. $P_r(s)=\dfrac{s}{N-(N-1)s}\cdot\dfrac{2s}{N-(N-2)s}\cdots\dfrac{rs}{N-(N-r)s}$.

15. $\mathbf{S}_r$ 是具有共同的几何分布的 r 个独立随机变量之和。因此

$$P_r(s)=\left(\frac{q}{1-ps}\right)^r,\qquad p_{r,k}=q^rp^k\binom{r+k-1}{k}.$$

16. $$\mathbf{P}\{\mathbf{R}=r\}=\sum_{k=0}^{\nu-1}\mathbf{P}\{\mathbf{S}_{r-1}=k\}\,\mathbf{P}\{\mathbf{X}_r\geqslant\nu-k\}$$

$$=\sum_{k=0}^{\nu-1}q^{r-1}p^k\binom{r+k-2}{k}p^{\nu-k}=p^\nu q^{r-1}\binom{r+\nu-2}{\nu-1}.$$

$$\mathbf{E}(\mathbf{R}) = 1 + \frac{qv}{p}, \quad \operatorname{var}(\mathbf{R}) = \frac{vq}{p^2}.$$

21. $u_n = q^n + \sum_{k=3}^{n} \binom{k-1}{2} p^3 q^{k-3} u_{n-k}$, 其中 $u_0 = 1$, $u_1 = q$, $u_2 = q^2$, $u_3 = p^3 + q^3$. 利用递推关系具有褶积形式这一事实,得

$$U(s) = \frac{1}{1-qs} + \frac{(ps)^3}{(1-qs)^3} U(s).$$

22. $u_n = pw_{n-1} + qu_{n-1}$, $v_n = pu_{n-1} + qv_{n-1}$, $w_n = pv_{n-1} + qw_{n-1}$. 因此 $U(s) - 1 = psW(s) + qsU(s)$; $V(s) = psU(s) + qs \cdot V(s)$; $W(s) = psV(s) + qsW(s)$.

第 十 三 章

1. 只需证明对于 $F(s) = 1$ 的所有根 $s \neq 1$ 我们有 $|s| \geqslant 1$ 且仅有周期情况才可能有 $|s| = 1$.

2. $u_{2n} = \left\{\binom{2n}{n} 2^{-2n}\right\}^r \sim (\pi n)^{-\frac{1}{2}r}$. 因此仅当 $r = 2$ 时 $\mathscr{E}$ 是常返的. 对于 $r = 3$ 数值积分的切线法给出

$$\sum_{n=1}^{\infty} u_{2n} \sim \pi^{-3/2} \int_{1/2}^{\infty} x^{-3/2}\, dx = \left(\frac{2}{\pi}\right)^{3/2} \approx \frac{1}{2}.$$

因此由(3.5) $\mathscr{E}$ 迟早要出现的概率近似为 $x = \frac{1}{3}$. 和的一个更精确的估计为 0.47, 因而得出 $x = 0.32$.

3. 每当 $n = k(a + b)$ 时 $\mathbf{S}_n = 0$ 是可能的,二项分布表明,对于这样的 n 我们有 $\mathbf{P}\{\mathbf{S}_n = 0\} \sim (a + b)^{1/2} (2\pi abk)^{1/2}$. 级数发散.

4. 由 $\Sigma f_k + \mathbf{P}\{\mathbf{X}_1 > 0\} \leqslant 1$ 可以断定,除 $\mathbf{P}\{\mathbf{X}_1 > 0\} = 0$ 的情况外,恒有 $f < 1$. 在这种情况,所有的 $\mathbf{X}_k \leqslant 0$ 且 $\mathscr{E}$ 或者在第一次试验中出现或者永不出现.

5. 设 $u_1, \cdots, u_N$ 为已给且对于 $n \geqslant N$, $u_n = p_1 u_{n-1} + p_2 u_{n-2} + \cdots + p_N u_{n-N}$. 则

$$\lim u_n = \frac{u_1 p_N + u_2(p_N + p_{N-1}) + \cdots + u_N(p_1 + p_2 + \cdots + p_N)}{p_1 + 2p_2 + 3p_3 + \cdots}.$$

如果 $u_k = N^{-1}$, 则 $\lim u_n = N^{-1}(u_1 + 2u_2 + \cdots + Nu_N)$.

6. $U_1(s) = \frac{1}{2} + \frac{1}{2} U(s)$, $F_1 = \sum_{k=1}^{\infty} \left(\frac{1}{2} F\right)^k = s^{-2} F^2(s)$.

7. $F_2(s)=\frac{1}{2}s\{1+F_1(s)\}=s^{-1}F(s)$.

8. $U_2(s)=\{1-F_2(s)\}^{-1}=\frac{1}{2}+\frac{1}{2}(1+s)(1-s^2)^{-1/2}$. 这表明，在时刻 $2n$ 初过一个正点的概率等于 1/2 即 $S_{2n}=0$ 的概率。

9. (a) $F(s)=qs(1-qs^r)^{-1}$, $\mu=1+rpq^{-1}$, $\sigma^2=r^2pq^{-2}$; (b) $\mathbf{Z}_n=n-\mathbf{N}_n$, $\mathbf{E}(\mathbf{Z}_n)\sim npq(q+pr)^{-1}$, $\operatorname{var}(\mathbf{Z}_n)=nr^2pq(q+pr)^{-2}$.

10. $U(s)=1+qs+\cdots+q^{r-1}s^{r-1}+q^rs^r(1-s)^{-1}$, $\mu^{-1}=q^r$.

11. $\mathbf{N}_n^*\approx(\mathbf{N}_n-714.3)/22.75$; $\Phi\left(\frac{2}{3}\right)-\Phi\left(-\frac{2}{3}\right)\approx\frac{1}{2}$.

12. $r_n=r_{n-1}-\frac{1}{4}r_{n-2}+\frac{1}{8}r_{n-3}$ 其中 $r_0=r_1=r_2=1$; $R(s)=(8+2s^2)(8-8s+2s^2-s^3)^{-1}$; $r_n\sim1.444248(1.139680)^{-n-1}$.

14. 如果 a_n 为一个长为 r 的 A 连贯在第 n 次试验出现的概率，则在 (7.5) 中以 α 代替 p，以 $1-\alpha$ 代替 q 即得 $A(s)$。设 $B(s)$ 与 $C(s)$ 分别为对应于 B 连贯与 C 连贯的函数.所需的母函数为 $F(s)=1-U^{-1}(s)$，其中在情况 (a) 中 $U(s)=A(s)$; 在 (b) 中 $U(s)=A(s)+B(s)-1$; 在 (c) 中 $U(s)=A(s)+B(s)+C(s)-2$。

15. 利用例 (8.b) 与问题 14 中的方法的一个简单的结合。

17. $u_n=Np$, $v_k(\infty)=Npq^k$.

19. 注意 $1-F(s)=(1-s)Q(s)$ 及 $\mu-Q(s)=(1-s)R(s)$，由此 $Q(1)=\mu$, $2R(1)=\sigma^2-\mu+\mu^2$. 幂级数 $Q^{-1}(s)=\Sigma(u_n-u_{n-1})s^n$ 当 $s=1$ 时收敛。

第 十 四 章

1. 输光概率仍然由(2.4)给出，其中 $p=\alpha(1-r)^{-1}$, $q=\beta(1-r)^{-1}$. 赌博的期望持续时间为 $\mathbf{D}_z(1-r)^{-1}$，其中 $\mathbf{D}_z$ 由 (3.4) 或 (3.5) 给出。

2. 边界条件(2.2)被 $q_0-\delta q_1=1-\delta$, $q_a=0$ 来代替。对应于(2.4)的解为

$$q_z=\{(q/p)^a-(q/p)^z\}(1-\delta)\div\{(q/p)^a(1-\delta)+\delta q/p-1\}.$$

边界条件 (3.2) 成为 $\mathbf{D}_0=\delta\mathbf{D}_1$, $\mathbf{D}_a=0$。

3. $q_z=pq_{z+2}+qq_{z-1}$ 对应于 (2.1)，且如果 $\lambda=p\lambda^3+q$ 即 $\lambda=1$ 或 $\lambda^2+\lambda=qp^{-1}$，则 $q_z=\lambda^z$ 是一个特解。输光概率为

$$q_x = \begin{cases} 1, & \text{如果 } q \geqslant 2p, \\ \left\{\left(\frac{1}{4}+\frac{q}{p}\right)^{1/2} - \frac{1}{2}\right\}^x, & \text{如果 } q \leqslant 2p. \end{cases}$$

5. $w_{z,n+1}(x) = pw_{z+1,n}(x) + qw_{z-1,n}(x)$，边界条件为 (1) $w_{0,n}(x) = w_{a,n}(x) = 0 (n \geqslant 1)$；(2) 对于 $z \neq x$ 有 $w_{z,0}(x) = 0$ 且 $w_{x,0}(x) = 1$。

6. 用 $w_{0,n}(x) = w_{1,n}(x)$ 及 $w_{a,n}(x) = w_{a-1,n}(x)$ 代替 (1)。

10. $\mathbf{P}\{\mathbf{M}_n < z\} = \sum_{x=1}^{\infty} (v_{x-z,n} - v_{x+z,n})$ $\mathbf{P}\{\mathbf{M}_n = z\} = \mathbf{P}\{\mathbf{M}_n < z+1\} - \mathbf{P}\{\mathbf{M}_n < z\}$。

11. 初过 x 必须在时刻 $k \leqslant n$ 出现，且质点在其后 $n-k$ 步由 x 回到 x。

第 十 五 章

1. P 的各行为 $(p, q, 0, 0)$, $(0, 0, p, q)$, $(p, q, 0, 0)$ 及 $(0, 0, p, q)$。对于 $n > 1$ 各行为 (p^2, pq, pq, q^2)。

2. (a) 链是不可约的和周期的；对于所有的 j, k, $p_{jk}^{(n)} \to 1/3$。(注意：P 是二重随机矩阵。) (b) 链具有周期 3，G 包含 E_1 与 E_2；状态 E_4 构成 G_2，E_3 构成 G_3。我们有 $u_1 = u_2 = 1/2$, $u_3 = u_4 = 1$。(c) 状态 E_1 与 E_3 构成一个闭集 S_1，E_4 与 E_5 构成另一个闭集 S_2，E_2 是非常返的。对应于闭集的矩阵是以 1/2 为元素的 2×2 矩阵。因此，如果 E_j 与 E_k 属于同一 S_r，则 $p_{jk}^{(n)} \to 1/2$；$p_{j2}^{(n)} \to 0$；最后，当 $k = 1, 3$ 时，$p_{2k}^{(n)} \to 1/2$，当 $k = 2, 4, 5$ 时，$p_{2k}^{(n)} \to 0$。(d) 链的周期为 3。令 $a = (0, 0, 0, 1/3, 1/3, 1/3)$, $b = (1, 0, 0, 0, 0, 0)$, $c = (0, 1/2, 1/2, 0, 0, 0)$，我们发现，$P^2 = P^5 = \cdots$ 的各行为 a, b, b, c, c, c，$P^3 = P^6 = \cdots$ 的各行为 b, c, c, a, a, a，$P = P^4 = \cdots$ 的各行为 c, a, a, b, b, b。

3. $p_{jj}^{(n)} = (j/6)^n$, $p_{jk}^{(n)} = (k/6)^n - ((k-1)/6)^n$ (当 $k > j$)，$p_{jk}^{(n)} = 0$ (当 $k < j$)。

4. $x_k = (3/4, 1/2, 1/4, 1/2)$, $y_k = (1/4, 1/2, 3/4, 1/2)$。

6. 令 $\mu = \Sigma n p_n$。如果 $\mu = \infty$，则状态是零状态。平稳分布：$u_k = \mu^{-1}(p_k + p_{k+1} + \cdots)$。

7. 当 $\Sigma(1-q_0)(1-q_1)\cdots(1-q_{n-1}) < \infty$ 时是遍历的。平稳分布与级数的项成比例。

8. $u_r = (p/q)^r (q-p)/p$。

9. $u_r = \{1 - p/q\}(p/q)^{r-1} \div \{1 - (p/q)^a\}$。

10. $p_{jj}=2j(N-j)/N^2$, $p_{j,j+1}=(N-j)^2/N^2$, $p_{j,j-1}=j^2/N^2$, $u_k=\binom{N}{k}^2\div\binom{2N}{N}$.

13. $$P=\begin{bmatrix} q & p & & & \cdots & 0 & 0 \\ 0 & 0 & 1 & & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ q & p & 0 & 0 & & 0 & 0 \end{bmatrix}.$$

14. 注意矩阵是二重随机的;利用例(6.b)。

15. 令 $p_{k,k+1}=1(k=1,\cdots,N-1)$, $p_{Nk}=p_k$.

16. $\Sigma u_j p_{jk}=u_k$, 于是 $U(s)=u_0(1-s)\{P(s)-s\}^{-1}$. 遍历性的充要条件是 $P'(1)<1$.

23. 设 M 为 x_j 的极大值。考虑状态 E_r, 其中 $x_r=M$.

26. 如果 $N\geqslant m-2$, 则变量 $\mathbf{X}^{(m)}$ 与 $\mathbf{X}^{(n)}$ 是独立的，因此矩阵 $p_{jk}^{(m,n)}$ 的三行与 $\mathbf{X}^{(n)}$ 的分布 (1/4, 1/2, 1/4) 恒等. 对于 $n=m+1$, 三行为 (1/2, 1/2, 0), (1/4, 1/2, 1/4), (0, 1/2, 1/2).

第 十 七 章

3. $\mathbf{E}(\mathbf{X})=ie^{\lambda t}$; $\mathrm{var}(\mathbf{X})=ie^{\lambda t}(e^{\lambda t}-1)$.

4. $P_n'=-\lambda nP_n+\lambda(n+1)P_{n+1}$.

$$P_n=\binom{i}{n}e^{-i\lambda t}(e^{\lambda}-1)^{i-nt}\quad(n\leqslant i).$$

$$\mathbf{E}(\mathbf{X})=ie^{-\lambda t};\ \mathrm{var}(\mathbf{X})=ie^{-\lambda t}(1-e^{-\lambda t}).$$

5. $P_n'(t)=-(\lambda+n\mu)P_n(t)+\lambda P_{n-1}(t)+(n+1)\mu P_{n+1}(t)$ $(n\leqslant N-1)$ 及 $P_N'(t)=-N\mu P_N(t)+\lambda P_{N-1}(t)$.

6. 具有 $\lambda_n=\lambda$ 及 $\mu_n=\mu$ 的生灭过程。

19. 解线微分方程的标准方法引出一组线性方程。参看第 417 页脚注中的提示。

参考文献

[1] A. Khintchine, Sur la des grands nombres, *Comptes rendus de l'Académie des Sciences*, 189 (1929), 477—479.

[2] J. W. Lindeberg, Eine neue Herleitum des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung, *Mathematische Zeitschrift*, 15 (1922), 211—225.

[3] W. Feller, Note on the large numbers and "fair" games, *Annals of Mathematical Statistics*, 16 (1945), 301—304.

[4] W. Feller, *Acta Scientiarum Litterarum Univ.* Szeged, 8 (1937), 191—201.

[5] W. Feller, Über den Zentralen Grenzwertsatz der Wahrscheilichkeitsrechnung, *Mathematische Zeitschrift*, 40 (1935), 521—559.

[6] V. Gončarov, Du domaine d'analyse Combinatoire, *Bulletin de l'Académie Sciences URSS, Sér. Math.* (用俄文写的，有法文摘要), 8 (1944), 3—48.

[7] W. Feller, The fundamental limit theorems in probability, *Bulletin of the American mathematical Society*, 51 (1945), 800—832.

[8] O. Bottema and S. C. Van Veen, Kansberekningen bij het biljartspel, *Nieuw Archief Voor Wiskunde* (荷兰文), 22 (1943), 16—33, 123—158.

[9] *University of California Publications in Statistics*, 1 (1952).

[10] D. G. Catcheside, Genetic effects of radiations, *Advances in Genetics*, edited by M. Demerec, 2, Academic Press, New York, 1948, 271—358, in particular 339.

[11] D. G. Kendall, On some modes of population growth leading to R. A. Fisher's logarithmic series distribution, *Bimetrika*, 35 (1948), 6—15.

[12] E. Schroedinger, Probability problems in nuclear chemistry, *Proceedings of the Royal Irish Academy*, 51, sect. A, No. 1 (December 1945).

[13] A. Lotka, Théorie analytique des associations biologiques, 2, *Actualités scientifiques et industrielles*, No. 780 (1939), 123—136, Hermann et Cie, Paris.

[14] R. A. Fisher, The genetical theory of natural Selection, Oxford, 1930, 73.

[15] D. G. Kendall, Stochastic processes and population growth, *Journal Royal Statistical Society*, 11 (1949), 230—265.

[16] T. E. Harris, Branching processes, *Annals of Mathematical Statistics*, 19 (1948), 474—494,

[17] C. Domb, On the use of a random parameter in combinatorial problems, *Proceedings Royal Philosophical Society*, Sec. A., **65** (1952), 305—309.
[18] H. Maier-Leibnitz, Die Koinzidenzmethode und ihre Anwendung auf Kernphysikalische Probleme, *Physikalische Zeitschrift*, **43** (1942), 333—362.
[19] P. Erdös, W. Feller, and H. Pollard, A theorem on power Series, *Bulletin of the American Mathematical Society*, **55** (1949), 201—204.
[20] D. Blackwell, Extension of a renewal theorem, *Pacific Journal of Mathematics*, **3** (1953), 315—320.
[21] W. Feller, Fluctuation theory of recurrent event, *Transactions of the American Mathematical Society*, **67** (1949), 98—119.
[22] *Cambridge Mathematical Journal*, **4** (1844), or The Mathematical and Other Writings of R. E. Ellis, Cambridge and London, 1863.
[23] L. Bachelier, Calcul des probabilités, Paris, 1912.
[24] G. Polya, Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Strassennetz, *Mathematische Annalen*, **84** (1921), 149—160.
[25] W. H. McCrea and F. J. W. Whipple, Random paths in two and three dimensions, *Proceedings of the Royal Society of Edinburgh*, **60** (1940), 281—298.
[26] W. Bartky, Multiple Sampling with Constant Probability, *Annals of Mathematical Statistics*, **14** (1943), 363—377.
[27] A. Wald. On Cumulative sum of random variables, *Annals of Mathematical Statistics*, **15** (1944), 283—296.
[28] A. Wald, Sequential analysis, John Wiley & Sons, New York, 1947.
[29] P. and T. Ehrenfest, Über zwei bekannete Einwände gegen das Boltzmannsche H-Theorem, *Physikalische Zeitschrift*, **8** (1907), 311—314.
[30] Ming Chen Wang and G. E. Uhlenbeck, On the theory of the Brownian motion II, *Reviews of Modern Physics*, **17** (1945), 323—342.
[31] M. Kac, Random walk and the theory of Brownian motion, *American Mathematical Monthly*, **54** (1947), 369—391.
[32] B. Friedman, A Simple urn model, *Communications On Pure and Applied Mathematics*, **2** (1949), 59—70.
[33] G. Malécot, Sur un problème de probabilités en chaine que pose la génetique, *Comptes rendus de l'Académie des Sciences*, **219** (1944), 379—381.
[34] W. Feller, Statistical aspect of ESP, *Journal of Parapsychology*, **4** (1940), 271—319.
[35] R. Fürth, Schwankungserscheinungen in der Physik, Sammlung Vieweg, Braunschweig, 1920. 17ff. 原来的观察发表于 *Physikalische Zeitschrift*, **19** (1918) and **20** (1919).

[36] A. Kolmogoroff, Zur Theorie der Markoffschen Ketten, *Mathematische Annalen,* **112** (1935), 155—160.

[37] W. Doeblin, Sur les propriétés asymptotigues de mouvements **régis** par certains types de Chaines Simples, Thesis, Paris, 1937.

[38] M. Fréchet, *Recherches théoriques modernes sur le calcul des probabilités,* **2** (théorie des événements en chaine dans le cas d'un nombre fini d'états possibles), Paris, 1938.

[39] B. Hostinsky, *Méthodes générales du calcul des probabilités,* fasc. 52 of the *Mémorial des Sciences mathématiques,* Paris, 1931.

[40] A. Kolmogorov, Anfangsgründe der Theorie der Markoffschen Ketten mit unendlich vielen möglichen Zuständen, *Matematičeskii Sbornik,* N. S., **1** (1936), 607—610. 这篇论文没有证明. 完全的解释仅用俄文在 *Bulletin de l'Université d' État à Moscou, Sect.* A. **1** (1937), 1—15 中给出.

[41] W. Doeblin, Sur deux problèmes de M. Kolmogoroff Concernant les Chaines dénombrables, *Bulletin Société Mathématique de France,* **66** (1939), 1—11.

[42] J. L. Doob, Topics in the theory of Markoff Chains, and also Markoff Chains — denumerable Case, *Transactions American Mathematical Society,* **52** (1942), 37—64, and **58** (1945), 455—473.

[43] K. L. Chung, Contributions to the theory of Markov Chains I, *Journal of Research,* National Bureau of Standards, **50** (1953), 203—208, and II, *Transactions American Mathematical Society,* **76** (1954), 397—419.

[44] C. Derman, A solution to a set of fundamental equations in Markov Chains, *Proceedings American Mathematical Society,* **5** (1954), 332—334.

[45] W. Feller, Boundaries induced by positive matrices, *Transactions American Mathematical Society,* **83** (1956), 19—54.

[46] T. E. Harris, On chains of infinite order, *Pacific Journal of Mathematics,* **5** (1955), Supplement 1, 707—724.

[47] S. Chandrasekhar, Stochastic problems in physics and astronomy, *Review of Modern Physics,* **15** (1943), 1—89, in particular p. 45.

[48] K. Yosida and Kakutani, Markoff processes with an enumerable infinite number of possible states, *Japanese Journal of Mathematics,* **16** (1939), 47—55.

[49] M. Kac, Random walk and the theory of Brownian motion, *American Mathematical Monthly,* **54** (1947), 369—391.

[50] G. Udny Yule, A mathematical theory of evolution.

[51] Dr. J. C. Willis, F. R. S., *Philosophical Transactions of the Royal Society,* London, Series B, **213** (1924), 21—87.

[52] A. G. M'Kendrick, Applications of Mathematics to medical problems, *Proceedings Edinburgh Mathematical Society,* **44** (1925), 1—34.

[53] W. Feller, Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in Wahrscheinlichkeitstheoretischer Behandlung, *Acta Biotheoretica*, 5 (1939), 11—40.

[54] W. H. Furry, On fluctuation phenomena in the passage of high-energy electrons through lead, *Physical Reviews*, 52 (1937), 569.

[55] W. Feller, On boundary Conditions for the Kolmogorov differential equations, *Annals of Mathematics*, 65 (1957), 527—570.

[56] W. Feller, The integrodifferential equations of Completely discontinuous Markov processes, *Transactions American Mathematical Society*, 48 (1940), 488—515.

[57] W. Ledermann and G. E. Reuter, Spectral theory for the differential equations of simple birth and death processes, *Philosophical Transactions Royal Society*, London, Ser. A, 246 (1954), 321—369.

[58] S. Karlin and J. McGregor, Representation of a class of stochastic processes, *Proceedings National Academy Sciences*, USA (6) 4(1955), 387—391.

[59] David G. Kendall, The generalized "birth and death" process, *Annals of Mathematical Statistics*, 19 (1948), 1—15.

[60] David G. Stochastic processes and population growth, *Journal of the Royal Statistical Society*, B, 11 (1949), 230—265.

[61] C. Palm, Intensitätsschwankungen im Fernsprechverkehr, *Ericsson Technics* (Stockholm), no. 44 (1943), 1—189, 特别可参阅57页.

[62] E. Brockmeyer, H. L. Halström, and Arne Jensen, The life and works of A. K. Erlang, *Transactions of the Danish Academy Technical Sciences*, No. 2, Copenhagen, 1948.

[63] T. C. Fry, Probability and its engineering uses, New York, 1928, 303ff.

[64] A. Kolmogoroff, Sur le problème d'attente, *Recueil Mathématique* [*Sbornik*], 38 (1931), 101—106.

[65] C. Palm, The distribution of repairmen in serving automatic machines (瑞典文), *Industritidningen Norden*, 75 (1947), 75—80, 90—94, 119—123.

[66] H. A. Adler and K. W. Miller, A new approach to probability problems in electrical engineering, *Transactions of the American Institute of Electrical Engineers*, 65 (1946), 630—632.

[67] A. Kolmogorov, Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung, *Mathematische Annalen*, 104 (1931), 415—458.

[68] J. L. Doob, Markoff chain — denumerable case, *Transactions American Mathematical Society*, 58(1945), 455—473.

[69] E. Vaulot, Delais d'attente des appels téléphoniques dans l'ordre inverse de leur arrivée, *Comptees Rendumptes*, Academie des Sciences, Paris, 238 (1954), 1118—1189.

[70] O. Lundberg, On random processes and their applications to sickness and accident statistics, Uppsala, 1940.

人 名 对 照 表

D. Ander 安德
L. Bachelier 巴切利尔
W. Bartky 巴特基
G. E. Bates 贝茨
D. Bernoulli 伯努利
D. Balackwell 伯拉克威尔
G. Cantor 康托
D. G. Chapman 查普曼
Chung K. L. 钟开莱
A. De Moivre 德模佛
C. Derman 德曼
W. Doeblin 多勃林
C. Domb 多姆
J. Doob 杜勃
P. and T. Ehrenfest 爱伦弗斯特
R. E. Ellis 艾利斯
P. Erdös 耳多斯
A. K. Erlang 埃尔兰
R. A. Fisher 费歇
M. Fréchet 弗雷谢
W. M. Furry 弗里
R. Fürth 弗思
F. Galton 格尔顿
T. E. Harris 哈里斯
M. Kac 卡斯
J. L. Lagrange 拉格朗日
P. S. Laplace 拉普拉斯
P. Lévy 列威
A. J. Lotka 洛特卡
A. G. M'kendrick 姆'肯德里克
G. Malécot 马兰考脱
R. Von Mises 封·米赛斯
J. Neyman 内曼
L. S. Ornstein 奥恩斯坦
C. Palm 帕姆
K. Pearson 皮尔逊
S. D. Poisson 普阿松
G. Polya 波利亚
G. E. Uhlenbeck 乌伦贝克
A. Wald 瓦尔德
F. J. W. Whipple 惠普尔
W. H. McCrea 麦克雷
N. Wiener 维纳
J. Wolfowitz 华尔夫维茨
S. Wright 赖特
G. U. Yule 尤尔
A. H. Колмогоров 柯尔莫果洛夫
A. M. Ляпунов 李雅普洛夫
A. A. Марков 马尔科夫
B. И. Романовский 罗曼诺夫斯基
A. Я. Хинчин 辛钦
П. Л. Чебышев 切比谢夫

索　　引

二画--四画

五　　画

六　　画

七 画

八 画

九 画

十 画

十 一 画

十 二 画

十二画以上